

DR EMBL: AF133086: AAF00109.1: 1
 DR HSP: P00763: IDPO.
 DR MEROPS: S01.302: 1
 DR InterPro: IPR000859: CUB.
 DR InterPro: IPR001314: Chymotrypsin.
 DR InterPro: IPR002172: LDL-recept-A.
 DR InterPro: IPR001254: Trypsin.
 DR Pfam: PF00431: CUB. 2.
 DR Pfam: PF00057: Ldl-recept-A. 4.
 DR Pfam: PF00089: trypsin. 1.
 DR PRINTS: PR00261: LDLRECEPTOR.
 DR PRINTS: PR00722: CHYMOTRYPSIN.
 DR SMART: SM00042: CUB. 2.
 DR SMART: SM00192: Ldla. 3.
 DR SMART: SM00020: Tryp-spc. 1.
 DR PROSITE: PS01180: CUB. 2.
 DR PROSITE: PS01209: LDLRA-1. 2.
 DR PROSITE: PS00068: LDLRA-2. 4.
 DR PROSITE: PS0240: TRYPsin-DOM. 1.
 DR PROSITE: PS00134: TRYPsin-HIS. 1.
 DR PROSITE: PS00135: TRYPsin-SER. 1.
 DR Signal-anchor: Glycoprotein; Hydrolase; Serine protease;
 KW Transmembrane; Repeat.
 FT DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 56 76 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 77 855 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 214 334 CUB. 1.
 FT DOMAIN 340 447 CUB. 2.
 FT DOMAIN 452 487 LDL-RECEPTOR CLASS A 1.
 FT DOMAIN 487 524 LDL-RECEPTOR CLASS A 2.
 FT DOMAIN 524 560 LDL-RECEPTOR CLASS A 3.
 FT DOMAIN 566 603 LDL-RECEPTOR CLASS A 4.
 FT DOMAIN 615 854 SERINE PROTEASE.
 FT ACT_SITE 656 656 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 711 711 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 805 805 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 485 485 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 855 AA: 94760 MM: 26143132C01F99C9 CRC64;

Query Match 99.9% score 4676; DB 1; Length 855;
 Best Local Similarity 99.9% Pred. No. 0;
 Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGSDARAKGGCGKPGFAGALAKYNSRHEKNGLEEGVEFLPVNNKVKVEKHGPRMVLAA 60
 1 MGSDARAKGGCGKPGFAGALAKYNSRHEKNGLEEGVEFLPVNNKVKVEKHGPRMVLAA 60
 QY 61 VLIGLLLVLLIGIGFLVWHLQYRDVRYQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
 61 VLIGLLLVLLIGIGFLVWHLQYRDVRYQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
 DB 61 VLIGLLLVLLIGIGFLVWHLQYRDVRYQKVFNGYMRITNENFVDAYENSNSTEFVSLASKV 120
 QY 121 KDALKLLYSGVPLGPGYHKEASVATFSEGSVALYVWSEFSIPQHLVEEAERVMAERVM 180
 121 KDALKLLYSGVPLGPGYHKEASVATFSEGSVALYVWSEFSIPQHLVEEAERVMAERVM 180
 DB 121 KDALKLLYSGVPLGPGYHKEASVATFSEGSVALYVWSEFSIPQHLVEEAERVMAERVM 180
 QY 181 LPPRASLSKSFVYTSVAAPFDKSTVORODNSCFGILHARGVELMRTTPGFPDPSYPA 240
 181 LPPRASLSKSFVYTSVAAPFDKSTVORODNSCFGILHARGVELMRTTPGFPDPSYPA 240
 DB 181 LPPRASLSKSFVYTSVAAPFDKSTVORODNSCFGILHARGVELMRTTPGFPDPSYPA 240
 QY 241 HARCQNALRGDADSVLSTFRSFDLASCDERGSDLVTVYNTLSMPDHALVOLCGTYPPS 300
 241 HARCQNALRGDADSVLSTFRSFDLASCDERGSDLVTVYNTLSMPDHALVOLCGTYPPS 300
 DB 241 HARCQNALRGDADSVLSTFRSFDLASCDERGSDLVTVYNTLSMPDHALVOLCGTYPPS 300
 QY 301 YNLTFSQNVLLITLITNTERRHPEAFEFQOLPRMSSCGGLRKAQGFNSPYPGHY 360
 301 YNLTFSQNVLLITLITNTERRHPEAFEFQOLPRMSSCGGLRKAQGFNSPYPGHY 360
 DB 301 YNLTFSQNVLLITLITNTERRHPEAFEFQOLPRMSSCGGLRKAQGFNSPYPGHY 360
 361 PNIDICTWNIIEVPNNQHVKKFKEFFYLLEPGVAGTCPRKDYVEINGEKKYCGERSQFVTS 420

DB 361 PNIDICTWNIIEVPNNQHVKKFKEFFYLLEPGVAGTCPRKDYVEINGEKKYCGERSQFVTS 420
 QY 421 NSKKIVRFHSDSYSDTGTGLAEYLSYSDSPCGGFTCTGRCIRRELKDCMADCTDH 480
 421 NSKKIVRFHSDSYSDTGTGLAEYLSYSDSPCGGFTCTGRCIRRELKDCMADCTDH 480
 DB 421 NSKKIVRFHSDSYSDTGTGLAEYLSYSDSPCGGFTCTGRCIRRELKDCMADCTDH 480
 QY 481 SDELNSCDAGHOFTCKNKKCKPLFWWCDSVNDGDNDEGSCSPAOPTFRCNGKLSK 540
 481 SDELNSCDAGHOFTCKNKKCKPLFWWCDSVNDGDNDEGSCSPAOPTFRCNGKLSK 540
 DB 481 SDELNSCDAGHOFTCKNKKCKPLFWWCDSVNDGDNDEGSCSPAOPTFRCNGKLSK 540
 QY 541 SQCCNGKDCGSDSDEASCPKVVNVYCTKHTYRCLNGLCLSKNPECDCGKEDCSGSDER 600
 541 SQCCNGKDCGSDSDEASCPKVVNVYCTKHTYRCLNGLCLSKNPECDCGKEDCSGSDER 600
 DB 541 SQCCNGKDCGSDSDEASCPKVVNVYCTKHTYRCLNGLCLSKNPECDCGKEDCSGSDER 600
 QY 601 DCCGGLRSTFROARVYGGTDADGEMPMOVSIALGCGHICGASLSPNMLVSAHCYID 660
 601 DCCGGLRSTFROARVYGGTDADGEMPMOVSIALGCGHICGASLSPNMLVSAHCYID 660
 DB 601 DCCGGLRSTFROARVYGGTDADGEMPMOVSIALGCGHICGASLSPNMLVSAHCYID 660
 QY 661 DRGFRYSDPTQMTAFGLADOSORSAFGVERRLKRIISHPFNDFTFYDIALLELEK 720
 661 DRGFRYSDPTQMTAFGLADOSORSAFGVERRLKRIISHPFNDFTFYDIALLELEK 720
 DB 661 DRGFRYSDPTQMTAFGLADOSORSAFGVERRLKRIISHPFNDFTFYDIALLELEK 720
 QY 721 AEVSSVVRPILCLPDASHVPAGKAIWVTGNGHTOYGGTGALLIQKEIRYINOTTCENL 780
 721 AEVSSVVRPILCLPDASHVPAGKAIWVTGNGHTOYGGTGALLIQKEIRYINOTTCENL 780
 DB 721 AEVSSVVRPILCLPDASHVPAGKAIWVTGNGHTOYGGTGALLIQKEIRYINOTTCENL 780
 QY 781 POOTPRMKGVFLSGVDSCGCGSGGLSSVEADGRIFOAGVYVWMDGCAORRKPVT 840
 781 POOTPRMKGVFLSGVDSCGCGSGGLSSVEADGRIFOAGVYVWMDGCAORRKPVT 840
 DB 781 POOTPRMKGVFLSGVDSCGCGSGGLSSVEADGRIFOAGVYVWMDGCAORRKPVT 840
 QY 841 RLPLFROMIKENTGV 855
 841 RLPLFROMIKENTGV 855
 DB 841 RLPLFROMIKENTGV 855

RESULT

ALIGNMENTS

```

RESULT 1
ST14_HUMAN
ID ST14_HUMAN STANDARD: PRT: 855 AA.
AC Q9Y5Y6;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-
DE TYPE SERINE PROTEASE 1) (MT-SPI).
GN ST14 OR PRSS14 OR SNC19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99303581; PubMed-10373424;
RA Lin C.Y., Anders J., Johnson M., Sang O.A., Dickson R.B.;
RT "Molecular cloning of cDNA for matriptase, a matrix-degrading serine
RT protease with trypsin-like activity.";
RL J. Biol. Chem. 274:18231-18236(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-99432178; PubMed-10500122;
RA Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
RT dissect complex biological processes and identify a membrane-type
RT serine protease in epithelial cancer and normal tissue.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN [3]
RP CHARACTERIZATION.
RC TISSUE=Milk;
RX MEDLINE-99303582; PubMed-10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matriptase
RT and a Kunitz-type serine protease inhibitor from human milk.";
RL J. Biol. Chem. 274:18237-18242(1999).
CC -!- FUNCTION: DEGRADES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
CC IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
CC ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
CC OR LYS AS THE P1 SITE.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF118224; AAD42765.2; -

```

Agencore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

CM protein - protein search, using sw model

Run on: January 20, 2002, 09:34:49 Search time 14.81 Seconds
(without alignments)
2116.707 Million cell updates/sec

Title: US-09-421-213-2
Perfect score: 4681
Sequence: 1 MGSDBARKGGCGPKDFGAGL.....PGVYTRLPFLFRMDIKENTGV 855

Scoring table: BLOSUM62
Gapop 10.0, Capext 0.5

Se ed: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008
Listing first 45 summaries

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4676	99.9	855	1	ST14_HUMAN
2	3901	83.3	855	1	ST14_MOUSE
3	727	15.5	1034	1	ENTK_PIG
4	712.5	15.2	1035	1	ENTK_BOVIN
5	682.5	14.8	1042	1	COR1_HUMAN
6	676.5	14.6	1019	1	ENTK_HUMAN
7	663.5	14.5	1069	1	ENTK_MOUSE
8	600	12.8	1113	1	COR1_MOUSE
9	586.5	12.5	704	1	CRAR_MOUSE
10	577	12.3	689	1	CRAR_HUMAN
11	560.5	12.0	492	1	TMS2_HUMAN
12	533.5	11.4	638	1	KAL_MOUSE
13	533	11.1	454	1	TMS3_HUMAN
14	518	11.1	603	1	CEAI_MOUSE
15	518	11.1	604	1	CEAI_RAT
16	514.5	11.0	638	1	KAL_RAT
17	513	11.0	435	1	TMS5_MOUSE
18	509.5	10.9	790	1	PLMN_PIG
19	501.5	10.7	786	1	STUB_DROME
20	500	10.7	638	1	KAL_HUMAN
21	496	10.6	638	1	KAL_MOUSE
22	496	10.6	810	1	PLMN_MOUSE
23	496	10.6	810	1	PLMN_HUMAN
24	495	10.6	437	1	TMS5_HUMAN
25	492.5	10.5	583	1	CEAI_HUMAN
26	491.5	10.5	812	1	PLMN_MOUSE
27	491	10.5	812	1	FA9_BOVIN
28	490.5	10.5	812	1	PLMN_BOVIN
29	490	10.5	417	1	HEPS_HUMAN
30	486	10.4	810	1	PLMN_HUMAN
31	484.5	10.4	343	1	PSS8_HUMAN
32	482	10.3	686	1	MAS2_HUMAN
33	480.5	10.3	416	1	HEPS_RAT

ALIGNMENTS

RESULT	1	ST14_HUMAN	STANDARD:	PRT:	855 AA.
34	480	10.3	625	1	FA1_HUMAN
35	479.5	10.2	461	1	FA9_HUMAN
36	479	10.2	343	1	PLMN_SHEEP
37	478	10.2	810	1	PLMN_ERIEU
38	477.5	10.2	342	1	PSS8_RAT
39	472.5	10.1	2616	1	NDL_DROME
40	472	10.1	275	1	TRY3_HUMAN
41	471.5	10.1	452	1	FA9_CANFA
42	467	10.0	275	1	TRY4_HUMAN
43	467	10.0	562	1	TPA_HUMAN
44	463	9.9	437	1	TMS4_HUMAN
45	461.5	9.9	267	1	TRY7_ANOGA

RESULT 1
ID ST14_HUMAN STANDARD: PRT: 855 AA.
AC Q95Y76;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (MATRIPTASE) (MEMBRANE-
TYPE SERINE PROTEASE 1) (MT-SPI).
GN ST14 OR PRSS14 OR SMC19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RN [2]
RN [3]
RP MEDLINE:99303582; PubMed-10373425;
RX Takeuchi T., Shuman M.A., Craik C.S.;
RT "Reverse biochemistry: Use of macromolecular protease inhibitors to
dissect complex biological processes and identify a membrane-type
serine protease in epithelial cancer and normal tissue.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:11054-11061(1999).
RN [3]
RP CHARACTERIZATION.
RC TISSUE-Milk;
RC MEDLINE:99303582; PubMed-10373425;
RA Lin C.Y., Anders J., Johnson M., Dickson R.B.;
RT "Purification and characterization of a complex containing matrilysin
and a kunitz-type serine protease inhibitor from human milk.";
J. Biol. Chem. 274:18237-18242(1999).
CC - FUNCTION: DEGRADATES EXTRACELLULAR MATRIX. PROPOSED TO PLAY A ROLE
IN BREAST CANCER INVASION AND METASTASIS. EXHIBITS TRYPSIN-LIKE
ACTIVITY AS DEFINED BY CLEAVAGE OF SYNTHETIC SUBSTRATES WITH ARG
OR LYS AS THE P1 SITE.
CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
CC - SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC - SIMILARITY: CONTAINS 2 CUB DOMAINS.
CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPsin FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF118224; AAD42765.2;

EMBL: AF133086; AAF00109.1;
 DR HSP: P00763; IDPO.
 DR MEROPS: S01.302.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LdL_recept_A.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00057; ldl_recept_a; 4.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00192; LDLA; 3.
 DR SMART: SM0020; TRYP_SPC; 1.
 DR PROSITE: PS01180; CUB; 2; 1.
 DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS00068; LDLRA_2; 4.
 DR PROSITE: PS00240; TRYP_SIN_DOM; 1.
 DR PROSITE: PS00134; TRYP_SIN_HIS; 1.
 DR PROSITE: PS00135; TRYP_SIN_SER; 1.
 KW Signal-anchor: Glycoprotein; Hydrolase; Serine protease;
 KW ansmembrane: Repeat.
 FT MAIN 1 55
 FT TRANSMEM 56 76
 FT DOMAIN 77 855
 FT DOMAIN 214 334
 FT DOMAIN 340 447
 FT DOMAIN 452 487
 FT DOMAIN 487 524
 FT DOMAIN 524 560
 FT DOMAIN 566 603
 FT DOMAIN 615 854
 FT ACT_SITE 656 656
 FT ACT_SITE 711 711
 FT ACT_SITE 805 805
 FT CARBOHYD 109 109
 FT CARBOHYD 302 302
 FT CARBOHYD 485 485
 FT CARBOHYD 772 772
 SQ SEQUENCE 855 AA: 94760 MW: 261431320.1 P99C9 CRC64;
 Query Match 99.98; Score 4676; DB 1; Length 855;
 Best Local Similarity 99.98; Pct. No. 0;
 Matches 854; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 MGSDRARKGGGKDFGAGLKYNSRHERKVNGLGVEFLPVNNVKKVKGHPGRWVLA 60
 Db 1 MGSDRARKGGGKDFGAGLKYNSRHERKVNGLGVEFLPVNNVKKVKGHPGRWVLA 60
 Qy 61 VLIGLLVLLGIGLFWHLQYRDVQVFNQYMRITNENFDAYNSNSTEFVSLASKV 120
 Db 61 VLIGLLVLLGIGLFWHLQYRDVQVFNQYMRITNENFDAYNSNSTEFVSLASKV 120
 Qy 121 KDALKLLVSGVPFLGPHYHESAVTAFSGSVIAYVWSEFSIPQHLVEAEAEVMAEERV 180
 Db 121 KDALKLLVSGVPFLGPHYHESAVTAFSGSVIAYVWSEFSIPQHLVEAEAEVMAEERV 180
 Qy 181 LPPRARSLSKFWTSVAVAFDSDKTQVORTDONSFGSLHARGVELMRTTTPGFDPSPY 240
 Db 181 LPPRARSLSKFWTSVAVAFDSDKTQVORTDONSFGSLHARGVELMRTTTPGFDPSPY 240
 Qy 241 HARCOWALRGDADSVLSITFSFDSACDERSGDLTVYNTLSPNEPHALVOLCGTTPPS 300
 Db 241 HARCOWALRGDADSVLSITFSFDSACDERSGDLTVYNTLSPNEPHALVOLCGTTPPS 300
 Qy 301 YNLTFHSSONVLLTLTINTERRHPGFEATFQPLPRSSCCGRLRKAOGTNSPYPGHY 360
 Db 301 YNLTFHSSONVLLTLTINTERRHPGFEATFQPLPRSSCCGRLRKAOGTNSPYPGHY 360
 Qy 361 PPNIDCTWNIEVPNNQHVVFSEKFFYLEPGVPAGTCPKDYVEINGEKYCGERSQFV 420

Db 361 PPNIDCTWNIEVPNNQHVVFSEKFFYLEPGVPAGTCPKDYVEINGEKYCGERSQFV 420
 Qy 421 NSNKITVRFHSDQSYDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
 Db 421 NSNKITVRFHSDQSYDTGFLAEYLSYDSSDPCPGQFTCTGRCIRKELRCGWDGADCTDH 480
 Qy 481 SDELNCSCDAGHOFCKKNKFKPLFWVCDSDVNDGDSDEQGCCSPAOTFRCNSGKCLSK 540
 Db 481 SDELNCSCDAGHOFCKKNKFKPLFWVCDSDVNDGDSDEQGCCSPAOTFRCNSGKCLSK 540
 Qy 541 SOQCNKDDCGDGSDEASCPKVVVYCTKHYRCLNGLCLSKGNPECDGKEDCSDGSEK 600
 Db 541 SOQCNKDDCGDGSDEASCPKVVVYCTKHYRCLNGLCLSKGNPECDGKEDCSDGSEK 600
 Qy 601 DCDGLRSFTROARVVGTDADGEMWQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
 Db 601 DCDGLRSFTROARVVGTDADGEMWQVSLHALGQGHICGASLISPNLWLSAAHCYID 660
 Qy 661 DRGFRYSDPTQWTAFLGLHDQSORSAPGVQBRRLKRIISHPFNFDTFYDIALLELEK 720
 Db 661 DRGFRYSDPTQWTAFLGLHDQSORSAPGVQBRRLKRIISHPFNFDTFYDIALLELEK 720
 Qy 721 AEYSSMRPVCICLPDASHVFPAGKAIWVTGWGTYGGTGALILQKGEIRVINOTTENLL 780
 Db 721 AEYSSMRPVCICLPDASHVFPAGKAIWVTGWGTYGGTGALILQKGEIRVINOTTENLL 780
 Qy 781 PQQITPRMCMVGLSGGVDSCQGGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKP 840
 Db 781 PQQITPRMCMVGLSGGVDSCQGGSGPLSSVEADGRIFQAGVSWGDCGCAQRNKP 840
 Qy 841 RLPLFRDWIKENTGV 855
 Db 841 RLPLFRDWIKENTGV 855
 RESULT 2
 ST14_MOUSE
 ID AC ST14_MOUSE STANDARD; PRT; 855 AA.
 AC P36677;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SUPPRESSOR OF TUMORIGENICITY 14 (EC 3.4.21.-) (EPITHIN).
 GN ST14 OR PRSS14.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C.B.17SCID; TISSUE=Thymus;
 RX MEDLINE=99216440; Pubmed=10199918;
 RA Kim M.G., Chen C., Lyu M.S., Cho E.G., Park D., Kozak C.,
 RA Schwartz R.H.;
 RT "Cloning and chromosomal mapping of a gene isolated from thymic
 RT stromal cells encoding a new mouse type II membrane serine protease,
 RT epithin, containing four LDL receptor modules and two CUB domains.";
 RL Immunogenetics 49:420-428(1999).
 RN [2]
 RP REVISTONS TO 23; 321; 325; 343; 409-410 AND C-TERMINUS.
 RC STRAIN-C.B.17SCID; TISSUE=Thymus;
 RX Kim M.G., Chen C., Cho E.G., Park D., Schwartz R.H.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Breast tumor;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN INTESTINE, KIDNEY, LUNG,
 CC AND THYMUS. NOT EXPRESSED IN SKELETAL MUSCLE, LIVER, HEART,
 CC TESTIS, AND BRAIN.

Db	121	KEALKLLYNEYPVLGPIYHAKSAVTAIFSEGSVIATYNSEFSLPHRLAEVEVRANAAYEVVT	1808
QY	181	LPFRARSLKSFVVTYSVAFAFPDSTKYVORTDONSFCFLHARGVELMFRTPGEPDPSYPA	240
Db	181	LPFRARSLKSFVLTLSVAFAFPIDPMLDORTQDNOSCFALHHAAGAIVTFRTPGFNPNSYPA	240
QY	241	HARCQMALRGDADSVLSLTFERSFDLASCDERGSDLVTYVNTLSMPHALVOLCGTYPPS	3000
QY	301	YNLFTHSSQNVLLTTLTNTERRRHPGEATFFOLPRHNSCGGRLRKQGFNSPYRGHY	3600
Db	301	YNLFTHSSQNVLLTTLTNTDRRHPGEATFFOLPKHNSCGGFLSDQGFNSPYRGHY	3600
QY	361	PNPIDCTWNIETVEPNHQHVYSFKFFYLLEGCPVAGTCPPKDYVELNGEKYCGERSQFVYTS	4200
Db	361	PNPINCNCWNIKVPPNNRVYKFAKFLYLVDPNVVPVSGCTKQYVELNGEKYCGERSQFVYSS	4200
QY	421	NSNKTIVRFHSDSYVTDGFLAEYLSYDSSDPCCPGQFTGRTGRCIRKELRCGDACDTH	4800
Db	421	NSSKITVHFHSDHYTDGTGFLAEYLSYDSDNDCPGMFCMKTGRCIRKELRCGDACDPDY	4800
QY	481	SDELNCSGDAGHOTCTCKNKKCKPLFWVCDSYNDCGDNSSDQSGSCPQOTFRCSNGKLSK	5400
Db	481	SDEHYCRCMATHQCTCKNQCFCKPLFWVCDSYNDCGDSDEGSCSPAGSKCSNGKCLPQ	5400
QY	541	SOQNCNGDDCGDGSDEASCPVNVVYCTKHTYRCLNLGLCSKGNPECDGKEDCSDSDEK	6000
Db	541	SOQKNGNDNGDGSDEASCSQSVNVYCTKHTYRQONLCLSKGNPECDGKEDCSDSDEK	6000
QY	601	DCDCGLSFTRQARVAVGCTADDECEMPQVYSLHALGGCHIGASLSLSPMLVSAHACYID	6600
Db	601	NCDCGLSFTRQAAVAVGCTNDADEEMQVYSLHALGGCHIGASLSLSPMLVSAHACFQD	6600
QY	661	DGRGRYSDPPQWTFELFLHDSQSSAGCVQERLKRITISHPEFNDFFEDDIALLELXP	7200
Db	661	DKNKYSYDTWTAFELGLDQSKRSASGAVDELKRTITHPSEFNDFFEDDIALELEKS	7200
QY	721	AEYSNMRPCLLPASHVFPAGKAIWYTGMGHTQYGGTGAIILOKGEIRYINOTTCENL	7800
Db	721	VEYSTVVRPCLLPATVHFPAGKAIWYTGMGHRTGEGTGALIILOKGEIRYINOTTCEDLM	7800
QY	781	PQQTTPRMKCVFLSGVDSCQDGGSLPSVEADGRIFQAGVSWMGCGCAQRNKPQVYT	8400
Db	781	PQQTTPRMKCVFLSGVDSCQDGGSLPSAEKDHMPDAGVSWMGCGCAQRNKPQVYT	8400
QY	841	RLPLFRDVIKENTGV 855	
Db	841	RLPLFRDVIKENTGV 855	

J. Biol. Chem. 269:19976-19982(1994).

-1- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN, PROCARBOXYPEPTIDASES, AND PROELASTASES.

-1- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN TRYPSINOGEN.

-1- SUBUNIT: HETEROPOLYMER OF A CATALYTIC (LIGHT) CHAIN, A MULTIDOMAIN (HEAVY) CHAIN, AND A MINI CHAIN.

-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).

-1- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS CLEAVED BY A TRYPSIN-LIKE PROTEASE.

-1- PTM: THE MINI CHAIN MAY BE CLEAVED BY ELASTASE.

-1- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.

-1- SIMILARITY: CONTAINS 1 SEA DOMAIN.

-1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.

-1- SIMILARITY: CONTAINS 1 MAM DOMAIN.

-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

is SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

EMBL: D30799; BAA06459.1; HSSP: P00763; LDPO.

MEROPS: S01.156.

InterPro: IPR000859; CUB.

InterPro: IPR001314; Chymotrypsin.

InterPro: IPR002172; LDL_recept_A.

InterPro: IPR000998; MAM.

InterPro: IPR000082; SEA.

InterPro: IPR001190; SRCR.

InterPro: IPR001254; Trypsin.

Pfam: PF00431; CUB; 2.

Pfam: PF00057; ldl_recept_a; 2.

Pfam: PF00629; MAM; 1.

Pfam: PF01390; SEA; 1.

Pfam: PF00530; SRCR; 1.

Pfam: PF00089; trypsin; 1.

PRINTS: PR00722; CHYMOTRYPSIN.

SMART: SM00042; CUB; 2.

SMART: SM00192; LDLa; 2.

SMART: SM00137; MAM; 1.

SMART: SM00200; SEA; 1.

SMART: SM00202; SR; 1.

SMART: SM00020; Tryp_Spc; 1.

PROSITE: PS01180; CUB; 2.

PROSITE: PS01209; LDLRA_1; 2.

PROSITE: PS00068; LDLRA_2; 2.

PROSITE: PS00740; MAM_1; 1.

PROSITE: PS50060; MAM_2; 1.

PROSITE: PS50024; SEA; 1.

PROSITE: PS00420; SRCR_1; FALSE_NEG.

PROSITE: PS50287; SRCR_2; 1.

PROSITE: PS00240; TRYPSIN_DOM; 1.

PROSITE: PS00134; TRYPSIN_HIS; 1.

PROSITE: PS00135; TRYPSIN_SER; 1.

Signal-anchor; Glycoproteina; Myristate; Hydrolase; Serine protease; Zymogen; Transmembrane; Repeat.

CHAIN 52 17

CHAIN 118 799

CHAIN 800 1034

TRANSMEM 19 47

SEA 52 169

DOMAIN 197 238

LDL-RECEPTOR CLASS A 1.

CUB 1.

DOMAIN	357	519	MAM.
FT DOMAIN	539	649	CUB 2.
FT DOMAIN	656	694	LDL-RECEPTOR CLASS A 2.
FT DOMAIN	693	786	SRCR.
FT DOMAIN	800	1034	SERINE PROTEASE.
FT ACT_SITE	840	840	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE	891	891	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE	986	986	MYRISTATE (POTENTIAL).
FT LIPID	2	2	BY SIMILARITY.
FT DISULFID	199	212	BY SIMILARITY.
FT DISULFID	206	225	BY SIMILARITY.
FT DISULFID	219	236	BY SIMILARITY.
FT DISULFID	558	670	BY SIMILARITY.
FT DISULFID	665	683	BY SIMILARITY.
FT DISULFID	677	692	INTERCHAIN (BY SIMILARITY).
FT DISULFID	787	911	BY SIMILARITY.
FT DISULFID	825	841	BY SIMILARITY.
FT DISULFID	925	992	BY SIMILARITY.
FT DISULFID	956	971	BY SIMILARITY.
FT DISULFID	982	1010	BY SIMILARITY.
FT CARBOHYD	116	116	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	147	147	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	170	170	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	194	194	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	283	283	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	343	343	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	350	350	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	403	403	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	455	455	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	485	485	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	518	518	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	549	549	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	645	645	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	697	697	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	701	701	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	721	721	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	740	740	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	761	761	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	804	804	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	863	863	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	902	902	N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD	964	964	N-LINKED (GLCNAC. .) (POTENTIAL).
SEQUENCE	1034 AA;	114776 MW;	0388C64CF64CC368 CRC64;

Query Match 15.5%; Score 727; DB 1; Length 1034;
 Best Local Similarity 23.8%; Pred. No. 1.4e-41;
 Matches 258; Conservative 135; Mismatches 330; Indels 360; Gaps 41;

QY	55	WVLA	AVLIGLLVLLGIGFLVHLQYRDVQKVF--NGYMRIT-----NENFVDAYEN	107
Db	21	FTAL	FAIMLVLCAGLIAVSWLTIKGSEKDAALGKSHEARGTMKITSVYTPNQLQKL--	78
QY	108	SNST	EFVSLASKVDALKLLYSGVPFLGYPYHKESAVTAFSEGSVIAYY-----W-SFEFS	160
Db	79	--SVDF	KVLAEDIQOMIGEIFQSSNLKNEY--KNSRVLOFENGSVIVIFDLLFAQWVSDE	135
QY	161	IPOH	LVBEAEERVAERVMV-----	180
Db	136	IKEL	IOGIEANKSQLVAFHIDVNSIDTESLENYSTTSPTTSDKLTSSPPATPCNV	195
QY	181	----LP	----PRARSLKSFV-----TSVVAFPTDSKTQVORTODNSCSFGLHARGVELMRFT	229
Db	196	SIEC	PGSRPCADALKCIAVLDFCDGELNCDPGSDSKICATACD-----GKFLLTES	249
QY	230	TPGP	PDSPYP---AHARCONALRGDADSVLSLTFRFDLASCDE-----GSDLV---	276
Db	250	SGSF	DAQAQPKLSEASVVCQMIIRVNOGLSTELNPSFYNTYSMDVNLNIEYGVGSSKILRA	309
QY	277	-----	-----TYNTLS-----	283
Db	310	SLWL	MNPCTIRIFSNQVTVTFLESDENDYIGFNATYATFAPNSTELNDEKINCNFEDGFC	369

QY 284PM 285
 DB 370 FWIQLDNDNEMERIGCTTPPTGPNFHTFGNAGFYISTPTGCGROERGLSLPL 429
 QY 286 EPHALVOLC.....GTPPSYN.....302
 DB 430 EP-TLEPVCLSPFYVMYGENVYKLSINISNDONIEKIIPOKEGVGNMNYGOVTLNETV 488
 QY 303 ---LTFISSONVLLITLITERRHGF-----EATFF-----QLPRSSCGG--RLR 345
 DB 489 EFVYAFNAFNKQFLSDYALDDISLTYGICNVSLYPTPTLPTSPPELP--TDCCGPELW 546
 QY 346 KAOGTFNSPYPPGYPHPTDCTNINIEVPNNOHVSKFYLEPVPAGCPADYVEI- 404
 DB 547 EPNTTFISMFNPNNYPOACVNNLNOKKN10LHEEEDLENIA-----DVEYIR 598
 QY 405 NGER-----YCGERSOPVYTSNSKRTIVRPHSDOSTYDTGFLAETLS---YDSSDPC 453
 DB 599 DGEEDSDLLAVYTGPGPVEDVFSTINRMVTLFTNDALRKGGKAFNTGYHLGTEPC 658
 QY 454 -PGOFTGTRCTRKELKRCGMDCTDHSDELNCSDAGHOFTCKNKFKCPLEWCDSVN 512
 DB 659 KEDNFOCENECVLLVNLCPFSHCKDGSDEACV-----RF-----LN 697
 QY 513 DCGDNDSEGCSCPAOTFE.....NGKCLSKSOOCNGKDDC---GDSDEASCP-----560
 DB 698 GTANNSG-----LVQFRTGSIWHTACAEWNTTQTSDDVCCOLLGLTGNSMPEFSSGG 750
 QY 561 ----KVVVVTCTHRYRCINGLCLSKGNPECDGKED-----GSDGSDKDCDGLRSFT 610
 DB 751 GPVFKLTATP-----NGSLILTASEQC--FEDSLILLOCHKS-----CGKKOVA 793
 QY 611 RQA--RVVGTDADEGWPMOVSLHALGCHIGCASLISPMVLVSAACHYIIDRGFRYS 668
 DB 794 QEVSPKIVGNDSEGAWPMVVALYNGQ--LLCGASLVSMDWLVSAHCVYCG---RNLE 848
 QY 669 PTQWTAFLGHDGOSRSPGVQERLKIISHPFNFTDYDALLELEKPAEYSSMR 728
 DB 849 PSMKKAALGHSMTSLNLSFQIVTRLADEIVINPHYNNRRRSDPAMHLEKXNYTVIQ 908
 QY 729 PICLPDASHVFPAGKATVYMGHTQYGTGALILLOKEIRVINQTCENLLPO-QITPR 787
 DB 909 PICLPENQVFPGRICISINGKVIYQSPADILQADVPLLSNENCOQOMEYNTEN 968
 QY 788 MKCVGFISSGVSDCGDSCGPLSSVEADGRIFOAGVSWKDCGAKRNKPCVYTRLPFRD 847
 DB 969 MMCAGYEEGIDSCGDSGCPMLCLE--NNRWMLAGVTSFGYOCALPMRPGVYARVPKETE 1027
 QY 848 WIK 850
 DB 1028 WIO 1030
 RESULT 4
 ENTK_BOVIN STANDARD; PRT; 1035 AA.
 AC P98072;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ENTEROPEPTIDASE PRECURSOR (EC 3.4.21.9) (ENTEROKINASE).
 GN PRSS7 OR ENTK
 OS Bos taurus (Bovine)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE-Duodenum;
 RA MEDLINE=94329561; PubMed=8052624;
 RX Kitamoto Y., Yuan X., Wu Q., McCoy D.W., Sadler J.E.,
 RT "Enterokinase, the initiator of intestinal digestion, is a mosaic

RT protease composed of a distinctive assortment of domains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).
 RN [2]
 RP SEQUENCE OF 801-1035 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=94043122; PubMed=8226855;
 RA Lavallie E.R., Rebentulla A., Racie L.A., Diblasio E.A.,
 RA Parezn C., Grant K.L., Light A., McCoy J.M.,
 RT "Cloning and functional expression of a cDNA encoding the catalytic
 RT subunit of bovine enterokinase.";
 RL J. Biol. Chem. 268:23311-23317(1993).
 RN [3]
 RP SEQUENCE OF 801-827.
 RC TISSUE-Intestine;
 RX MEDLINE=92189715; PubMed=1799406;
 RA Light A., Janska H.,
 RT "The amino-terminal sequence of the catalytic subunit of bovine
 RT enterokinase.";
 RL J. Protein Chem. 10:475-480(1991).
 CC - FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPsin, CHYMOTRYPsin AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPsinOGEN TO TRYPsin WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPsinOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES.
 CC - CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPsinOGEN.
 CC - SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.
 CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC - TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.
 CC - PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPsin-LIKE PROTEASE.
 CC - SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC - SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC - SIMILARITY: CONTAINS 1 SCR DOMAIN.
 CC - SIMILARITY: CONTAINS 1 SCR DOMAIN.
 CC - SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U09859; AAB40026.1;
 DR EMBL: L19663; AAA16035.1;
 DR PIR: A61436; A61436.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.156;
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR000998; MAM.
 DR InterPro: IPR000082; SEA.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00057; LDL_recept_a; 2.
 DR Pfam: PF00629; MAM; 1.
 DR Pfam: PF01390; SEA; 1.
 DR Pfam: PF00530; SRCR; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPsin.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00192; LDla; 2.
 DR SMART: SM00137; MAM; 1.
 DR SMART: SM00200; SEA; 1.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00020; TRYP_Spc; 1.
 DR PROSITE: PS0180; CUB; 2.

DR PROSITE: PS01209; LDLRA_1; 2.
 DR PROSITE: PS0068; LDLRA_2; 2.
 DR PROSITE: PS00740; MAM_1; 1.
 DR PROSITE: PS0060; MAM_2; 1.
 DR PROSITE: PS0024; SEA; 1.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS00287; SRCR_2; 1.
 DR PROSITE: PS00240; TRYPsin_DOM; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; 1.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 KW Signal-anchor: Glycoprotein; Myristate; Hydrolase.
 KW Signal-anchor: Transmembrane; Repeat; Alternative splicing.
 FT CHAIN 1 800
 FT CHAIN 801 1035
 FT TRANSMEM 19 47
 FT DOMAIN 54 169
 FT DOMAIN 197 238
 FT DOMAIN 240 350
 FT DOMAIN 358 520
 FT DOMAIN 540 650
 FT DOMAIN 657 695
 FT DOMAIN 694 787
 FT MAIN 801 1035
 FT .T_SITE 841 841
 FT ACT_SITE 892 892
 FT ACT_SITE 987 987
 FT LIPID 2
 FT DISULFID 199 212
 FT DISULFID 206 225
 FT DISULFID 219 236
 FT DISULFID 659 671
 FT DISULFID 666 684
 FT DISULFID 678 693
 FT DISULFID 788 912
 FT DISULFID 826 842
 FT DISULFID 926 933
 FT DISULFID 957 972
 FT DISULFID 983 1011
 FT CARBOHYD 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 170 170
 FT CARBOHYD 194 194
 FT CARBOHYD 233 233
 FT CARBOHYD 263 263
 FT CARBOHYD 264 264
 FT CARBOHYD 404 404
 FT CARBOHYD 456 456
 FT CARBOHYD 486 486
 FT CARBOHYD 519 519
 FT CARBOHYD 550 550
 FT CARBOHYD 646 646
 FT CARBOHYD 698 698
 FT CARBOHYD 722 722
 FT CARBOHYD 741 741
 FT CARBOHYD 762 762
 FT CARBOHYD 864 864
 FT CARBOHYD 903 903
 FT CARBOHYD 965 965
 FT VARSPIC 166 192
 FT CONFLICT 808 808
 SQ SEQUENCE 1035 AA; 114887 MW; E207970B08296E13 CRC64;

Query Match 15.2%; Score 712.5; DB 1; Length 1035;
 Best Local Similarity 23.3%; Pred. No. 1.3e-40;
 Matches 253; Conservative 149; Mismatches 319; Indels 365; Gaps 45;

Qy 55 WVLAALVIGLLVLGGLVHLYQYDRVRQKVF--NGYMR-ITNENFVDAYNSNST 111
 Db 21 FAVLPVILVALCAGLIAVSWLSIQGSKDAAFGKSHEARGTLKIISGATYNPHLQDKLSV 80
 Qy 112 EFVSLASKVKDAKLLYSGVPLGYPYHKESAVTAFSEGSVIAYY-----W-SFESIQQH 164

Db 81 DFKVLAFDIOQIDDIFOSSNLKNEY-KNSRVLOFENGSIIVIFDLDFQWSDKNVKEE 139
 Qy 165 LVEEAERVMAEERYVM-----LPPRARSLSKSFVVTSV-----V 197
 Db 140 LIQIGTEANKSOLVTFHIDLNSIDITASLENFSTISPATSEK-LTTSIPLATPGNVSI 197
 Qy 198 AFPDTSKVQFTQNSCSFGLHARGVELM-----RFTTGGF 233
 Db 198 ECPDPSRL-----CADALKCAIDLFCOGLNCPGSDENKTCATACDGRFLTGS 249
 Qy 234 PDS-----PYPAH---ARCQWALRGDADSVLSLTFRSFD----- 264
 Db 250 SGSEALHYKPSNNTSAVCWIIIRVQGLSIQLNFDFYFYADVNLNIEGMSKILR 309
 Qy 265 -----LASCOBERGSDLV---TVYNTLSPMEPHALVQL----- 293
 Db 310 ASLWSNPGIIRIFSNQVTATFLIQSDE--SDYIGFKVITYAFNSKELNNEYKINCNFED 367
 Qy 294 --C-----GTYPSPYNLTF-HS----- 307
 Db 368 GCFWIQDLNDDNEWERTQGSTFPSTGPTFDHTFGNESGYISTPTGPGRRRERVGLLT 427
 Qy 308 -----SONVLLITLITNTRRHHPGFEATFFQ----- 333
 Db 428 LPDPTPEOACLSPWYMYGENVYKLSINISDQ---NWEKTIQKEGNYQNNYQGV 484
 Qy 334 -----LPRMSGCG 342
 Db 485 LNETVEKVSFYGKNOILSDIALDDISLTGYICNVSVYPEPTLVPTPPPELP--TDCGG 542
 Qy 343 --RLRKAQCTENSYPYPCGHYPNIDCTWNIENPNQHVKVSFKFYLLPGLPGVAGTCKPD 400
 Db 543 PHDLWEPTTFTSFNPNSPYNAQFCIWNNAQKGNQLHFQEFLENIA-----D 594
 Qy 401 YVEI-NGEKYCGRSQFV-----VTSNSKIVTRFHSQSVTDTGFLAEYLS-- 446
 Db 595 VVEIRGE---GDSLSFLAVYTGPGVNDVFSTNRTVLEITDNMLAKQFKANFTGY 651
 Qy 447 -YDSDPG-PCQFTCTGRCIRKELRCGDWADCTHSDCLNC-----SCDAGH--OFTC 496
 Db 652 GLGIPCEKEDNFOCKDGEICPLVNLCDGPHCKDGSDEAHCVRFLNFTTOSGLVQFRI 711
 Qy 497 KNKECKPLFWCDSVNDGDSDEGSCCPAQTFRCSNGKCLSKSQCCNGKDCGDSDE 556
 Db 712 QS-----IWHV-----ACAEN-----WTTQISDDVC-----QLLG---LGTGNS 743
 Qy 557 ASC-----PKVNVVTCTKHYRCLNGLSKGNPECDGKEDCDGSDGDEKDCD---CGLR 607
 Db 744 VPTSTGGGPVNLNTP-----NGSLIILTPSOQC-----LEDLSILLQCNKSCGKK 791
 Qy 608 SFTROA---RVVGGTDADGEWPMVSLHALGQGHICGASLISPNMLVSAACHYIDDRGFR 665
 Db 792 LVTQEVSPKIVGGSDSREGANPWVVALYFDQ-QVCGASLVSRDLVSAACHVYG----R 846
 Qy 666 YSDPTQNTAFGLHDSQORSAPGVQERELKRIISHPPFNDETFDYDIALLELEKPAEYSS 725
 Db 847 NMEPSKAVJGLHMASNLTSQIETRLIDQIVINPHYNKRKNNDIAMHLEKVNITD 906
 Qy 726 MVRPCLPDASHVFPAGKAIWVTGNGHTQYGGTGALILQKGEIRVINOTTGENTLLPQ-QI 784
 Db 907 YIQPICLPEENQVPPGGRICSIAGNAGLIIYOGSTADVLQEADVPLLSNEKCOQMPFYNI 966
 Qy 785 TPRMVCVGLSGVDSCGDSGGLSSVYEDGRIFQAGVSWGDCGAORNKPGYVITRLP 844
 Db 967 TENMVCAGYEAGDVGSCGDSGGLMCOE--NNRWLLAGVTSFGYQCALPNRPGVYARVPR 1025
 Qy 845 FRDWIK 850
 Db 1026 FTEWIO 1031
 RESULT 5
 CORI_HUMAN

ID COR1 HUMAN STANDARD: PRT: 1042 AA.
 DT 09Y505: 09UH2: 40. Created
 DT 20-AUG-2001 (Rel. 40. Last sequence update)
 DT 20-AUG-2001 (Rel. 40. Last sequence update)
 DT 20-AUG-2001 (Rel. 40. Last sequence update)
 DE ATRIAL NATRIURETIC PEPTIDE-CONVERTING ENZYME (EC 3.4.21.1) (PRO-ANP-
 CONVERTING ENZYME) (CORIN) (HEART SPECIFIC SERINE PROTEINASE ATC2).
 GN CRN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Heart;
 RA MEDLINE:99262646; PubMed:10329693;
 RA Yan W., Sheng N., Seto M., Morser J., Wu Q.;
 RT Corin, a mosaic transmembrane serine protease encoded by a novel cDNA
 RT from human heart. J. Biol. Chem. 274:14926-14935(1999).
 RL J. Biol. Chem. 274:14926-14935(1999).
 RN [2]
 RP SEQUENCE OF 734-1040 FROM N.A.
 RC TISSUE-Heart;
 RA MEDLINE:20534769; PubMed:11082206;
 RA Hooper J.D., Scaram A.L., Clarke B.E., Normyle J.F., Antalis T.M.;
 RT Localization of the mosaic transmembrane serine protease corin to
 RT heart myocytes. J. Biol. Chem. 267:6931-6937(2000).
 RL Eur. J. Biochem. 267:6931-6937(2000).
 RN [3]
 RP CHARACTERIZATION.
 RA MEDLINE:20359740; PubMed:10880574;
 RA Yan W., Wu F., Morser J., Wu Q.;
 RT Corin, a transmembrane cardiac serine protease, acts as a pro-atrial
 RT natriuretic peptide-converting enzyme. J. Biol. Chem. 274:14926-14935(1999).
 RL Proc. Natl. Acad. Sci. U.S.A. 97:8525-8529(2000).
 CC -1- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY
 CC BETWEEN ARG-123 AND SER-124.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART. EXPRESSED IN HEART
 CC MYOCYTES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FRIZZLED (FZ) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC his SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: AF133845; AAD31850.1;
 CC EMBL: AF133248; AAF21966.1;
 CC HSSP: P01130; IAUJ.
 DR HMM: 605236; -
 DR InterPro: IPR000024; -
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_receptor_A.
 DR Pfam: PF01392; Fz_2.
 DR Pfam: PF00057; Idl_recept_a; 7.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00261; LDLRECEPTOR.
 DR SMART: SM00063; FRI; 2.
 DR SMART: SM00192; LDLa; 7.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00202; TRYP_SPC; 1.
 DR PROSITE: PS50038; Fz; 2.

DR PROSITE: PS01209; LDLA_1; 6.
 DR PROSITE: PS50068; LDLA_2; 7.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00135; TRYPSIN_HIS; FALSE-NEG.
 DR PROSITE: PS00420; SRCR_1; FALSE-NEG.
 DR PROSITE: PS50287; SRCR_2; FALSE-NEG.
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Repeat.
 FT DOMAIN 1 45
 FT TRANSMEM 46 66
 FT DOMAIN 134 259
 FT DOMAIN 268 304
 FT DOMAIN 305 340
 FT DOMAIN 341 377
 FT DOMAIN 378 415
 FT DOMAIN 450 573
 FT DOMAIN 579 614
 FT DOMAIN 615 653
 FT DOMAIN 654 690
 FT DOMAIN 690 786
 FT DOMAIN 802 1042
 FT ACT_SITE 843 843
 FT ACT_SITE 882 882
 FT ACT_SITE 985 985
 FT DISULFID 790 912
 FT DISULFID 828 844
 FT DISULFID 955 970
 FT DISULFID 981 1010
 FT CARBOHYD 80 80
 FT CARBOHYD 104 104
 FT CARBOHYD 135 135
 FT CARBOHYD 141 141
 FT CARBOHYD 231 231
 FT CARBOHYD 245 245
 FT CARBOHYD 251 251
 FT CARBOHYD 305 305
 FT CARBOHYD 320 320
 FT CARBOHYD 376 376
 FT CARBOHYD 413 413
 FT CARBOHYD 446 446
 FT CARBOHYD 451 451
 FT CARBOHYD 469 469
 FT CARBOHYD 567 567
 FT CARBOHYD 651 651
 FT CARBOHYD 697 697
 FT CARBOHYD 761 761
 FT CARBOHYD 1022 1022
 FT MUTAGEN 985 985
 FT CONFLICT 854 854
 FT CONFLICT 876 876
 SQ SEQUENCE 1042 AA; 116564 MW; 7705398BEB67AD2 CRC64;
 Query Match 14.8%; Score 692; DB 1; Length 1042;
 Best Local Similarity 26.8%; Pred. No. 3.2e-39;
 Matches 206; Conservative 95; Mismatches 246; Indels 222; Caps 31;
 OY 202 DSKYVORTDQNSGFGHARGVLMFTTGPDPSPAPAHAROMALRGDAD-----253
 DB 365 DHCCVYKSDSEVNC--CHSGGLVECR-----NGCCIPSTFGCD-----GDEDCDSDEE 412
 OY 254 --SVLSITFRSPD-----LASCDERGSDLVTYVNTL--SPMERHALVQLCGTYP-- 298
 DB 413 NCSVIDTSCGEGORCLYNCLDSCG--GSLCDPNLSLNCSOCPEITL-ELCMNLPYN 469
 OY 299 ----PSY-----NLTHSSQ-----NVLLITLTNTERRRHPPG 327
 DB 470 STSYPNVFGHRTQKASISWESSLPALVQTNCKYKLMFSCITLVKPCDVNGFERP-- 527
 OY 328 EATFFQLPHNUSCGGRLLKAKQGTENSPYRGHYPRNIDCT-WNIEVYNNQHVAVSKFEY 386

Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump A., Shillhabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E., Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P., Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;

"The DNA sequence of human chromosome 21.";

Nature 405:311-319(2000).

[4]

SEQUENCE OF 749-1019 FROM N.A.

RA RN

TISSUE=Duodenum;

RC

MEDLINE=94329561; PubMed=8052624;

RC

Kitamoto Y., Yuan Q., McCourt D.W., Sadler J.E.;

RC

"Enterokinase, the initiator of intestinal digestion, is a mosaic protease composed of a distinctive assortment of domains.";

RA

Proc. Natl. Acad. Sci. U.S.A. 91:7588-7592(1994).

RA

- !- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN, PROCARBOXYPEPTIDASES, AND PROELASTASES;

RA

- !- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-|-ILE-7 BOND IN TRYPSINOGEN.

RA

- !- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND.

RA

- !- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).

RA

- !- TISSUE SPECIFICITY: INTESTINAL BRUSH BORDER.

RA

- !- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS CLEAVED BY A TRYPSIN-LIKE PROTEASE.

RA

- !- DISEASE: DEFECTS IN PRSS7 CAUSE LIFE-THREATENING INTESTINAL MALABSORPTION CHARACTERIZED BY DIARRHEA AND FAILURE TO THRIVE.

RA

- !- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.

RA

- !- SIMILARITY: CONTAINS 2 CUB DOMAINS.

RA

- !- SIMILARITY: CONTAINS 1 SEA DOMAIN.

RA

- !- SIMILARITY: CONTAINS 1 SRCR DOMAIN.

RA

- !- SIMILARITY: CONTAINS 1 MAM DOMAIN.

RA

- !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY.

RA

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isu-sib.ch/announcement/> or send an email to license@sib-sib.ch).

RA

EMBL; U09860; AAC50138.1; -

RA

EMBL; Y19124; CAB65555.1; -

RA

EMBL; Y19125; CAB65555.1; JOINED.

RA

EMBL; Y19126; CAB65555.1; JOINED.

RA

EMBL; Y19127; CAB65555.1; JOINED.

RA

EMBL; Y19128; CAB65555.1; JOINED.

RA

EMBL; Y19129; CAB65555.1; JOINED.

RA

EMBL; Y19130; CAB65555.1; JOINED.

RA

EMBL; Y19131; CAB65555.1; JOINED.

RA

EMBL; Y19132; CAB65555.1; JOINED.

RA

EMBL; Y19133; CAB65555.1; JOINED.

RA

EMBL; Y19134; CAB65555.1; JOINED.

RA

EMBL; Y19135; CAB65555.1; JOINED.

RA

EMBL; Y19136; CAB65555.1; JOINED.

RA

EMBL; Y19137; CAB65555.1; JOINED.

RA

EMBL; Y19138; CAB65555.1; JOINED.

RA

EMBL; Y19139; CAB65555.1; JOINED.

RA

EMBL; Y19140; CAB65555.1; JOINED.

RA

EMBL; Y19141; CAB65555.1; JOINED.

RA

EMBL; Y19142; CAB65555.1; JOINED.

RA

EMBL; Y19143; CAB65555.1; JOINED.

RA

EMBL; AL163218; CAB90392.1; -

RA

EMBL; AL163217; CAB90389.1; -

RA

DR HSP: P00763; 1DPO.
 DR MEROPS; S01.156; .
 DR MIM; 226200; .
 DR InterPro; IPR000859; CUB.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR002172; LDL_recept_A.
 DR InterPro; IPR000998; MAM.
 DR InterPro; IPR000082; SEA.
 DR InterPro; IPR001150; SRCR.
 DR InterPro; IPR001254; TRYPsin.
 DR Pfam; PF000431; CUB; 2.
 DR Pfam; PF000057; ldl_recept_a; 2.
 DR Pfam; PF006629; MAM; 1.
 DR Pfam; PF01380; SEA; 1.
 DR Pfam; PF00530; SRCR; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; P000722; CHYMOTRYPSIN.
 DR MART; SM00042; CUB; 2.
 DR MART; SM00192; LDLa; 2.
 DR SMART; SM00137; MAM; 1.
 DR SMART; SM00200; SEA; 1.
 DR SMART; SM00202; SR; 1.
 DR SMART; SM00020; TRYP_Spc; 1.
 DR PROSITE; PS01180; CUB; 2.
 DR PROSITE; PS01209; LDLRA_1; 2.
 DR PROSITE; PS00740; MAM_1; 1.
 DR PROSITE; PS00068; LDLRA_2; 2.
 DR PROSITE; PS50060; MAM_2; 1.
 DR PROSITE; PS50024; SEA; 1.
 DR PROSITE; PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE; PS50287; SRCR_2; 1.
 DR PROSITE; PS50240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 DR Signal-anchor: Glycoprotein; Myristate; Hydrolase;
 KM Serine protease: Zymogen; Transmembrane; Repeat;
 CHAIN 1 784
 FT CHAIN 1 784
 FT TRANSMEM 19 1019
 FT DOMAIN 52 169
 FT DOMAIN 182 223
 FT DOMAIN 225 334
 FT DOMAIN 342 504
 FT DOMAIN 524 634
 FT DOMAIN 641 679
 FT DOMAIN 678 771
 FT OMAIN 785 1019
 FT ACT_SITE 825 825
 FT ACT_SITE 876 876
 FT ACT_SITE 971 971
 FT LIPID 2 2
 FT DISULFID 184 197
 FT DISULFID 191 210
 FT DISULFID 204 221
 FT DISULFID 643 655
 FT DISULFID 650 668
 FT DISULFID 662 677
 FT DISULFID 772 896
 FT DISULFID 810 826
 FT DISULFID 910 977
 FT DISULFID 941 956
 FT DISULFID 967 995
 FT CARBOHYD 116 116
 FT CARBOHYD 147 147
 FT CARBOHYD 179 179
 FT CARBOHYD 328 328
 FT CARBOHYD 335 335
 FT CARBOHYD 388 388
 FT CARBOHYD 440 440
 FT CARBOHYD 470 470
 FT CARBOHYD 503 503
 FT CARBOHYD 534 534
 FT CARBOHYD 630 630

FT CARBOHYD 682 682 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 706 706 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 725 725 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 848 848 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 887 887 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 909 909 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 949 949 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 134 134 O -> E (IN REF. 3).
 FT CONFLICT 732 732 S -> P (IN REF. 3).
 FT CONFLICT 754 771 SOOCLDSLRLQCNHKS -> RRNKNKIDALSPILILIA
 FT CONFLICT 754 771 (IN REF. 3).
 SO SEQUENCE 1019 AA; 112923 MM; BGAAA245f6DA4563 CRC64;
 Query Match 14.6%; Score 682.5; DB 1; Length 1019;
 Best Local Similarity 32.5%; Pred. No. 1.4e-38;
 Matches 80; Conservative 80; Mismatches 202; Indels 91; Gaps 20;
 QY 333 QLRPMSSCGG-RLKRAQGFNSPIYRGHPNIDCNWINEVNNQHVAKSEFFYLEP 390
 DB 519 ELP--TDCGGPFELMEPNTFSSSTFNPSYPLAFVCWMLNAQKNIOLEHFEFDLEN 575
 QY 391 GVPAGTCPRDYVEI-NGEK-----YCGERSQFVYTSNKNITVRFHSDQSYTOTGFL 441
 DB 576 -----INDVEIRDSGEADSLILAVYTPGPYKDYFTTNMTVLLITNDVLRAGFK 628
 QY 442 AEYLS---YDSDPCPC-PTCTGTCRCIRKELRQDGNADCTDHSDELNCSCDAGHC-- 494
 DB 629 ANFTTGVLHGLPECKADHQCKNGECVPLVNLCDGLHCEGDSDEADCV-----REFNG 683
 QY 495 TCNN---KCKPLFWYCDVNDGNSDEQSCSPAQOTRCSNGKCLSSQCNCKGDC 550
 DB 684 TTNNGLVRFRIOSIW---HTCAEN-----WTOISNDVC-----QLL 719
 QY 551 GDSDEASCKVNVVCTKHTYRCLNGCLSKGPECDG-----KEDCSDG-----D 598
 DB 720 GLSGNSKPIFTS-----DGGPFVKLTAPDGLILTPSOCLDLSIRLQCN 768
 QY 599 EKDCDCGLRSFTQARVYGTDADCEMPQVSLHALGQGHGASLISPMVLVAANCY 658
 DB 769 HKSCGKLAQODITPRKVGSSNAKEGAMPVVGIL-YGGRLLCGASLVSDDMLVSAHCV 827
 QY 659 IDRGFFYSPTQMTALGLHDSQSRASAPGOERLRKRITSHFPNDFTDYDIALLELE 718
 DB 828 YG---RNLPSKWTALGLHMSNLTSPQTPRLIDEIYINPYNRRRNDNDIAMHLE 883
 QY 719 KPAEYSMBPCLDPAHSYFPAGKAIWVGHTGYGTGALLILOKEIRVINOTTCEN 778
 DB 884 FKNVNTYIPICLPEBNQVFPFGNCSIAKGTVYOGTANILQADVPILSNERCQ 943
 QY 779 LLPO-QITPRMNCVGLSGVDSCQDGGSLSSVEADGRIFGAGVYVSWDGAQRNKP 837
 DB 944 OMEYNTNTEWMCAGYEGLIDSCQSGGSLPMQCE-NRMWFLAGVTSFQYKCALPVRPG 1002
 QY 838 VYTRPLPFRDWIK 850
 DB 1003 VYAVRSRTFEMIQ 1015
 RESULT 7
 ENTK_MOUSE STANDARD; PRT; 1069 AA.
 AC P97435;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ENTEROPEPTIDASE (EC 3.4.21.9) (ENTEROKINASE).
 GN PRS7 OR ENTK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE=Duoenum;
 RX MEDLINE=98147142; PubMed=9486188;
 RA Yuan X., Zheng X., Lu D., Rubin D.C., Pung C.Y.M., Sadler J.E.;
 RT "Structure of murine enterokinase (enteropeptidase) and expression in
 small intestine during development.";
 RL Am. J. Physiol. 274:G342-G349(1998).
 CC -!- FUNCTION: RESPONSIBLE FOR INITIATING ACTIVATION OF PANCREATIC
 CC PROTEOLYTIC PROENZYMES (TRYPSIN, CHYMOTRYPSIN AND CARBOXYPEPTIDASE
 CC A). IT CATALYZES THE CONVERSION OF TRYPSINOGEN TO TRYPSIN WHICH IN
 CC TURN ACTIVATES OTHER PROENZYMES INCLUDING CHYMOTRYPSINOGEN,
 CC PROCARBOXYPEPTIDASES, AND PROELASTASES (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: SELECTIVE CLEAVAGE OF 6-LYS-1-ILE-7 BOND IN
 CC TRYPSINOGEN.
 CC -!- SUBUNIT: HETERODIMER OF A CATALYTIC (LIGHT) CHAIN AND A
 CC MULTIDOMAIN (HEAVY) CHAIN LINKED BY A DISULFIDE BOND (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (PROBABLE).
 CC -!- PTM: THE CHAINS ARE DERIVED FROM A SINGLE PRECURSOR THAT IS
 CC CLEAVED BY A TRYPSIN-LIKE PROTEASE (BY SIMILARITY).
 CC -!- SIMILARITY: CONTAINS 2 LDL-RECEPTOR CLASS A DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 CUB DOMAINS.
 CC -!- SIMILARITY: CONTAINS 1 SEA DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U73378; AAB37317.1;
 CC HSP: P00763; LDPO.
 CC MEROPS: S01.156;
 CC MGD: MGI:1197523; Prss7.
 CC InterPro: IPR000859; CUB.
 CC InterPro: IPR001314; Chymotrypsin.
 CC InterPro: IPR002172; LDL_recept_A.
 CC InterPro: IPR000598; MAM.
 CC InterPro: IPR000082; SEA.
 CC InterPro: IPR001190; SRCR.
 CC InterPro: IPR001254; Trypsin.
 CC Pfam: PF00431; CUB; 2.
 CC Pfam: PF00057; ldl_recept_a; 2.
 CC Pfam: PF06629; MAM; 1.
 CC Pfam: PF01390; SEA; 1.
 CC Pfam: PF00530; SRCR; 1.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC SMART: SM00042; CUB; 2.
 CC SMART: SM00192; LDLa; 2.
 CC SMART: SM00137; MAM; 1.
 CC SMART: SM00200; SEA; 1.
 CC SMART: SM00202; SR; 1.
 CC SMART: SM00020; TRYP_SPQ; 1.
 CC PROSITE: PS01180; CUB; 2.
 CC PROSITE: PS01209; LDLA_1; 2.
 CC PROSITE: PS00068; LDLA_2; 2.
 CC PROSITE: PS00740; MAM_1; 1.
 CC PROSITE: PS00060; MAM_2; 1.
 CC PROSITE: PS00024; SEA; 1.
 CC PROSITE: PS00420; SRCR_1; FALSE_NEG.
 CC PROSITE: PS0287; SRCR_2; 1.
 CC PROSITE: PS0240; TRYPSIN_DOM; 1.
 CC PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.
 CC Signal-anchor: Glycoprotein; Myristate; Hydrolase;
 KW Signal-anchor: Zymogen; Transmembrane; Repeat.

CHAIN	1	829	NON-CATALYTIC CHAIN (HEAVY CHAIN).
FT CHAIN	830	1069	CATALYTIC CHAIN (LIGHT CHAIN).
FT TRANSMEM	19	47	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT DOMAIN	52	169	SEA.
FT DOMAIN	227	268	LDL-RECEPTOR CLASS A 1.
FT DOMAIN	270	379	CUB 1.
FT DOMAIN	387	549	MAM.
FT DOMAIN	569	679	CUB 2.
FT DOMAIN	723	816	LDL-RECEPTOR CLASS A 2.
FT DOMAIN	830	1069	SRCR.
FT ACT_SITE	874	874	SERINE PROTEASE.
FT ACT_SITE	925	925	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE	1021	1021	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT LIPID	2	2	MYRISTATE (POTENTIAL).
FT DISULFID	229	242	BY SIMILARITY.
FT DISULFID	236	255	BY SIMILARITY.
FT DISULFID	249	266	BY SIMILARITY.
FT DISULFID	688	700	BY SIMILARITY.
FT DISULFID	695	713	BY SIMILARITY.
FT DISULFID	707	722	BY SIMILARITY.
FT DISULFID	817	945	INTERCHAIN (BY SIMILARITY).
FT DISULFID	859	875	BY SIMILARITY.
FT DISULFID	959	1027	BY SIMILARITY.
FT DISULFID	991	1006	BY SIMILARITY.
FT DISULFID	1017	1045	BY SIMILARITY.
FT CARBOHYD	147	147	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	197	197	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	212	212	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	373	373	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	380	380	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	433	433	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	515	515	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	579	579	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	675	675	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	727	727	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	751	751	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	770	770	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	791	791	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	897	897	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	936	936	N-LINKED (GLCNAC.) (POTENTIAL).
FT CARBOHYD	999	999	N-LINKED (GLCNAC.) (POTENTIAL).
SQ SEQUENCE	1069 AA;	118735 MW;	E62549E463743C3D CRC64;

Query Match 14.5%; Score 676.5; DB 1; Length 1059;
 Best Local Similarity 29.2%; Pred. No. 3.7e-38;
 Matches 186; Conservative 93; Mismatches 220; Indels 137; Gaps 23;

Qy	258	LTFRSFDLASCDSRGSLVTVYNTLSPNEPHALVOLCTYPPSYNLTFHSSONVLLITLI	317
Db	522	VVFNAFRNRGCTTALDDISLTNGICSSQSPYPEPTLVPTPP	563
Qy	318	TINTERHPGFATFOLPRMSSCG--RLRAQGTNSPYPGHYPPNIDCTWNEVPNN	375
Db	564	-----ELP-----TDCGGPFELWEPNSTFSSPNFPDKYPNQASCIWLNNAORG	606
Qy	376	QHKVSPFMFFYLLEPGVPAGTCKPDYVEI--NGE-----KYCERSQFVVVTSNSKIT	426
Db	607	KNIOHFOEFPLEN-----INDVVEVRDGGEDFSDLLAVYTGPGVKOLFSTTNMT	658
Qy	427	VRFHSDQSYTDGFLAEYLS---YDSSDPC-PGQFTCTGTCIRKELRCGMACTDHS	482
Db	659	VIFTTNETRRKPKANFTSGYILGIPPCQDDEFOCKDGNICPLGNLCDSYPHCRDGS	718
Qy	483	ELNC-----SCDAGHOFTCKNFKCPFLWVCDSDVNDGCDNDEQCCSCPAQTRCSN	534
Db	719	EASCVREFLNGTRSNGLVQFNHS-----IWHI-----ACAEN-----WTTQISN	758
Qy	535	GKLSKSOQCKGKDDCGDGSDEACP-----KYNVVTCTKTYRCNLGLCLSKGN	584
Db	759	EVC-----HLLGLGSANSMPISSSTGGGPPVRVN-----QAPNGSLITPS	799

QY 585 PEGDGE-----DCSDGDEKDCDGLRSFTROA--RVVGSTDEDEGEMPOVSLHALSGG 638
 DB 800 LQCSQDLSLLQCNHKS-----CGEKNVTKQKSPKIVGSDQACAPVVALYHRDNR 853
 QY 639 H--ICGASLISFNMLYSAHACYIDRGRFYSDPTQMTAFGLHDSQSPAPCVQERRLK 695
 DB 854 TDRLLCGASLVSDMLVSAHACYR---RNLDPRTMTAVLGTHMSNLTSPOVYRKYD 909
 QY 696 RIISHPFDFDFTDIDALLELEKPAEYSSMWRTCLPDASHVFPAGKAIWVTGCHTQY 755
 DB 910 QIVINPHYDRRRKVDIAMHLEFKVNYTVDYDIPCLPEENOIFIFGRCSISMGVYDKI 969
 QY 756 -GGTALLI0KGEIRVINTTCENILPQ-QIIRRMVCVPLSGVSCGSDGSLSSVE 813
 DB 970 NAGSTVDVLEKADVLEKIPVSEKQQLPEYNITESMTCAGYEESGIDSCGSDGSLPQCE 1029
 QY 814 ADGRIFQAGVSMGDCGCAQRNPGVYTRPLPLFEDMI 849
 DB 030 -NNRMFLVGTVSGVGCALPHNPGVYRVRSQFIEMI 1064
 RESULT 8
 COR1.MOUSE STANDARD: PRT: 1113 AA.
 AC 092319:
 DT 20-AUG-2001 (rel. 40, Created)
 DT 20-AUG-2001 (rel. 40, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE ATRIAL NATRIURETIC PEPTIDE-CONVERTING ENZYME (EC 3.4.21.-) (PRO-ANP-
 DE CONVERTING ENZYME) (CORIN) (LOW DENSITY LIPOPROTEIN RECEPTOR RELATED
 DE PROTEIN 4).
 GN CRN OR LRPA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Granata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98429596; PubMed=9756624;
 RA Tomita Y., Kim D.-H., Magbori K., Fujino T., Yamamoto T.T.;
 RA "A novel low-density lipoprotein receptor-related protein with type II
 RT membrane protein-like structure is abundant in heart".
 RL J. Biochem. 124:784-789(1998).
 CC -1- FUNCTION: CONVERTS PRO-ANP TO ANP. CLEAVES PRO-ANP SPECIFICALLY
 CC BETWEEN ARG-122 AND SER-123 (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC -1- SIMILARITY: CONTAINS 7 LDL-RECEPTOR CLASS A DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FRIZZLED (FZ) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AB013874; BAA34373.1;
 DR HSSP: P01130; IAUJ.
 DR MGD: MG1:1349451; LP4.
 DR InterPro: IPR000024; Fz_domain.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR InterPro: IPR001314; Thymotrypsin.
 DR InterPro: IPR002172; LDL_recept_A.
 DR Pfam: PF01392; Fz; 2.
 DR Pfam: PF00057; Ldl_recept_A; 7.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PRO0722; CHYMOTRYPSIN.

DR PRINTS: SM00263; LDLRECEPTOR.
 DR SMART: SM00263; FRI; 2.
 DR SMART: SM00192; LDLA; 7.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00202; TRYP_Spc; 1.
 DR PROSITE: PS50038; FZLRA_1; 6.
 DR PROSITE: PS50109; LDLRA_2; 7.
 DR PROSITE: PS50068; LDLRA_2; 7.
 DR PROSITE: PS50420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS50287; SRCR_2; 1.
 DR PROSITE: PS50240; TRYPsin_DOM; 1.
 DR PROSITE: PS00134; TRYPsin_HIS; FALSE_NEG.
 DR PROSITE: PS00135; TRYPsin_SER; 1.
 KW Hydrolyase; Serine protease; Transmembrane; Signal-anchor;
 KW Glycoprotein; Repeat.
 FT DOMAIN 1
 FT TRANS 113 133
 FT FT
 FT DOMAIN 134 1113
 FT FT
 FT DOMAIN 201 327
 FT FT
 FT DOMAIN 336 372
 FT FT
 FT DOMAIN 373 408
 FT FT
 FT DOMAIN 409 445
 FT FT
 FT DOMAIN 446 483
 FT FT
 FT DOMAIN 518 641
 FT FT
 FT DOMAIN 647 682
 FT FT
 FT DOMAIN 683 721
 FT FT
 FT DOMAIN 722 757
 FT FT
 FT DOMAIN 758 853
 FT FT
 FT DOMAIN 869 1113
 FT FT
 FT ACT_SITE 910 910
 FT ACT_SITE 959 959
 FT ACT_SITE 1052 1052
 FT DISULFID 857 979
 FT DISULFID 895 911
 FT DISULFID 1022 1037
 FT DISULFID 1048 1077
 FT CARBOHYD 147 147
 FT CARBOHYD 202 202
 FT CARBOHYD 208 208
 FT CARBOHYD 298 298
 FT CARBOHYD 317 317
 FT CARBOHYD 373 373
 FT CARBOHYD 411 411
 FT CARBOHYD 444 444
 FT CARBOHYD 481 481
 FT CARBOHYD 519 519
 FT CARBOHYD 537 537
 FT CARBOHYD 635 635
 FT CARBOHYD 719 719
 FT CARBOHYD 765 765
 FT CARBOHYD 828 828
 FT CARBOHYD 970 970
 FT CARBOHYD 1089 1089
 FT SEQUENCE 1113 AA; 122984 MW; B845B2C5F200DB8C CRC64;

Query Match 14.2%; Score 663.5; DB:1; Length 1113;
 Best Local Similarity 32.7%; Pred. No. 2.9e-37;
 Matches 158; Conservative 67; Mismatches 187; Indels 71; Gaps 16;

QY 428 RIISDSYDTGTGLAEIYLDSSDPGQFTGTCRCIRKELCDGHCADCTHSDSLNCS 487
 DB 627 QFPEESSDNQTCLLP---NEDVEECSSHFKCRSGKCVLSGRCDQADDDSDSENCG 683
 QY 488 CDAGHGFCTK-NKECFPLFWCDVNDGNSDEGCS-CPAOTFRCSNCKCSKSOQC 545
 DB 684 CKERALMECFNNQCKLHTLCGFPDCCPSNDEKNSFCQDNLELCANHECVPRDLMD 743
 QY 546 GKDDCGDSCDEASCPYV-----VVTGKHYRCLNG-----LC----- 579
 DB 744 GAVDCSDSDSEMCGVLTLSKNGNSSILTVHKSAREHVCADGWRRLTSLQACQOMGLGEP 803

QY 580 -LSKNPECDGKE-----DCSDGSDKED-----C----DCGLRST 610
 DB 804 SVTKLIPGEGQWURLPNMNLKSTLQELLVYRHSCPSRSEISLLCSKQDCGRPPAA 863
 QY 611 R-QARVGGTDADECEWQVSLHALGOGHICGASLISPNMLVSAHGYIDDRGFYSDP 669
 DB 864 RNNKILGRTSRPGRWQWQSLQSESGHICGCVLIKKWLTVAHCF-EGR-----EDA 918
 QY 670 TOWTAFGLHDSQSAFQVQERRLRIISHPFENFTFDYDIALLELEKPAEYSSMYRP 729
 DB 919 DWKVVFGINLDPHG-FMOTREVKITILLHPYRAVVDYDISVVELSDDINETSYVRP 977
 QY 730 ICLDPASHVFPAGKAIWVTGHTGYGTGALILQKGEIRVINQTTENLIPQO-ITPRM 788
 DB 978 VCLP-EEYLEPDTYGYITGWHG--MGNMPFKLQEGEVRIPLQCCQSYFDMKTIITRM 1035
 QY 789 MCVGLSGVDSQCGSGPLSSVEADGRIFOAGVYVSGDGAOR-NKPGVYTRLPFLRD 847
 DB 1036 ICAGVESTVDSMGDSGGLPYCVRPGQWTLFGLTSMGSCVCFKVLGPGVYNSVYFVG 1095
 QY 848 WIK 850
 DB 196 WIE 1098

RESULT 9
 TMS2_MOUSE STANDARD: PRT: 490 AA.
 AC Q9JIO8: Q9JIK4; Q9QY82;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-) (EPITHELIALIN) (PLASMIC
 DE TRANSMEMBRANE PROTEIN %)
 CN TMPSR52
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN-BALB/C;
 RX MEDLINE-21104370; PubMed-11169526;
 RA Vaarala M.H., Porvari K.S.; Kellokumpu S., Kyllonen A.P., Viikko P.T.;
 RT "Expression of transmembrane serine protease TMPSR52 in mouse and
 RL human tissues";
 RL J. Pathol. 193:134-140(2001).
 RP [2]
 RP SEQUENCE FROM N.A.
 RA Han J., Kim S.;
 RT "Putative transmembrane protease x";
 RL .bmitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C;
 RA Jacquinet E.J., Rao N.V., Rao G.N., Hoidal J.R.;
 RT "A novel mosaic serine protease, epitheliasin";
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MAINLY IN PROSTATE AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

DR EMBL: AF199362; AAF97867.1; -
 DR EMBL: AF243500; AAF64186.1; -
 DR EMBL: AF113596; AAF21308.1; -
 DR HSSP: P00761; IAKS
 DR MGD: MGI:1354381; Tmprss2.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_receptin_A.
 DR InterPro: IPR001190; SRCR.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00057; ldl_receptin_a; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00192; LDL; 1.
 DR SMART: SM00202; SR; 1.
 DR SMART: SM00020; Tryp_SPC; 1.
 DR PROSITE: PS01209; LDLRA_1; 1.
 DR PROSITE: PS50068; LDLRA_2; 1.
 DR PROSITE: PS00420; SRCR_1; FALSE_NEG.
 DR PROSITE: PS00287; SRCR_2; 1.
 DR PROSITE: PS0240; TRYPSIN_DOM; 1.
 DR PROSITE: PS0134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 DR KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 FT DOMAIN 1 83 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 84 104 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN) (POTENTIAL).
 FT DOMAIN 105 490 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 111 149 LDL-RECEPTOR CLASS A.
 FT DOMAIN 150 242 -SRCR.
 FT DOMAIN 254 490 SERINE PROTEASE.
 FT ACT_SITE 294 294 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 343 343 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 439 439 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT SITE 253 254 CLEAVAGE (POTENTIAL).
 FT DISULFID 76 125 BY SIMILARITY.
 FT DISULFID 119 138 BY SIMILARITY.
 FT DISULFID 132 147 BY SIMILARITY.
 FT DISULFID 171 230 BY SIMILARITY.
 FT DISULFID 184 240 BY SIMILARITY.
 FT DISULFID 243 363 BY SIMILARITY.
 FT DISULFID 279 295 BY SIMILARITY.
 FT DISULFID 408 424 BY SIMILARITY.
 FT DISULFID 435 463 BY SIMILARITY.
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 474 474 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 122 122 S->L (IN REF. 3).
 FT CONFLICT 178 178 S->N (IN REF. 3).
 FT CONFLICT 320 320 Y->H (IN REF. 1).
 FT CONFLICT 474 474 N->D (IN REF. 1).
 SQ SEQUENCE 490 AA; 53479 MW; 07D2B03EA4D8A1A9 CRC64;

Query Match 12.88; Score 600; DB 1; Length 490;
 Best Local Similarity 34.38; Pred. No. 2.2e-33;
 Matches 137; Conservative 62; Mismatches 152; Indels 48; Gaps 11;
 QY 461 TGCIRKELRCIDGWADCTDHSDELNCSCDAGHQFTCKNKFKPLFW--VCDSDVNDGCGNS 518
 DB 122 SGTCSISLWCDGVAHCPNGEDENRCVLYCQSFILQVSSQRKAWYPCV--DWSESY 179
 QY 519 DEQCSCSPAQTFRCNSGKCLSKSQCCNGKDCGSDGSDASCPKVNVTCTKHTYRCLNGL 578
 DB 180 GRAAC-----KDMCYKNFYSSQGIPOSGATSPMKLVN----- 213
 QY 579 CLSKGNPECDKCDKSDGSDK-----DC-DCGLRSTFTRQARVYCGTDDAGEHPWQVSL 632
 DB 214 --SSGNVDLYKKLHSDSCSSSRMVVSLRCIECGVRSVRKQSRIVGGLNASPCDWPQVSL 271
 QY 633 HALQOGHICGASLISPNMLVSAHGYIDDRGFYSDPTQWTAFLGLHDSQSAAPGVQER 692
 DB 272 HVQGV-HVCGGSIITPEWIVTAAHCVPEP-----LSGPRYWTAFAGILRQSLMFYV--SRH 324

QY 693 RLKRIISHPEFDETFDYDIALLELEKFAEYSSWVRPCLPDASHVFPAGKAIWVGWH 752
 DB 325 QVEKVISHPNDSTKTKNDIALMLQTPLANRDLVPCPLPNCMDLDOECBHSNGA 384
 QY 753 TOYCGTALLIQKEIRVINOCTEEN--LLEPOITPRMVCFLISGVDSGCGSGLS 810
 DB 385 TYEKGTSDVLANAMVPLLEBCKNSKTYNNLTPAMICAGFLOGSDSCGSGPLV 444
 QY 811 SVEADGRIFONGVSMGDGCKQRNKKPGYTRLPLEFPMI.849
 DB 445 TLK-NGIMWLIGDTSMGSCAKALRPGYGVNTVETDWT 482
 RESULT 10
 CRAR_MOUSE STANDARD; PRT: 704 AA.
 AC P98064;
 DT 1-FEB-1996 (Rel. 33, Created)
 DT 1-FEB-1996 (Rel. 33, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE COMPLEMENT-ACTIVATING COMPONENT OF RA-REACTIVE FACTOR PRECURSOR
 DE (EC 3.4.21.-) (RA-REACTIVE FACTOR SERINE PROTEASE P100) (RAKF)
 DE (MANNAN-BINDING LECTIN SERINE PROTEASE 1)
 GN MASPI OR CRAR
 OS Mus musculus (Mouse)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN 11
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-BALB/C; TISSUE=Liver.
 RX MEDLINE=94179811; PubMed=8133044;
 RA Takayama Y., Takada F., Takahashi A., Kawakami M.;
 RA "A 100-kDa protein in the C4-activating component of Ra-reactive
 RT factor is a new serine protease having module organization similar to
 RT C1r and C1s."
 RT J. Immunol. 152:2308-2316(1994).
 RN 12
 RP SEQUENCE OF 465-704 FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN-BALB/C; TISSUE=Liver.
 RX MEDLINE=93176166; PubMed=8439319;
 RA Takahashi A., Takayama Y., Hattuse H., Kawakami M.;
 RA "Presence of a serine protease in the complement-activating component
 RT of the complement-dependent bactericidal factor, Rarf, in mouse
 RT Serum."
 RT Biochem. Biophys. Res. Commun. 190:681-687(1993).
 RC 1- FUNCTION: COMPONENT OF THE BACTERICIDAL RA-REACTIVE FACTOR RARF
 CC WHICH SPECIFICALLY BINDS TO RA AND R2 POLYSACCHARIDES EXPRESSED BY
 CC CERTAIN ENTEROBACTERIA. IT TRIGGERS THE ACTIVATION OF COMPLEMENT
 CC CASCADE BY ACTIVATING THE C4 AND C2 COMPONENTS. IT ACTIVATES THE
 CC C4 COMPONENT BY CLEAVING THE ALPHA-CHAIN OF C4.
 CC 1- SUBUNIT: RARF CONSISTS OF A COMPLEMENT-ACTIVATING COMPONENT
 CC (CRARF) AND A POLYSACCHARIDE-BINDING (MANNOSE-BINDING) COMPONENT.
 CC CRARF IS AN HETERODIMER OF A HEAVY (P70) AND A LIGHT CHAIN (29)
 CC LINKED BY A DISULFIDE BOND.
 CC 1- TISSUE SPECIFICITY: LIVER.
 CC 1- DOMAIN: CRARF HAS A MODULE ORGANIZATION SIMILAR TO C1R AND C1S.
 CC 1- SIMILARITY: CONTAINS 2 SUSHI (SCR) DOMAINS.
 CC 1- SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
 CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: D16492; BAA03944.1;
 CC HSSP: P00736; IAP0.

DR MEROPS: S01.198;
 DR MD: MG1.88492; Asx_hydroxyl.
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000859; CUB.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR000436; Sushi_SCR_CCP.
 DR InterPro: IPR001254; trypsin.
 DR Pfam: PF00431; CUB; 2.
 DR Pfam: PF00084; sushi; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00032; CCP; 2.
 DR SMART: SM00042; CUB; 2.
 DR SMART: SM00179; EGF_Ca; 1.
 DR SMART: SM00020; TRYP_SPE; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_Ca; 1.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Glycoprotein; Sushi; Repeat; Signal; EGF-like domain; Hydroxylation.
 FT SIGNAL 1 24
 FT CHAIN 25 704
 FT FT
 FT CHAIN 25 704
 FT CHAIN 25 704
 FT CHAIN 454 703
 FT DOMAIN 25 143
 FT DOMAIN 144 187
 FT DOMAIN 190 302
 FT DOMAIN 305 368
 FT DOMAIN 371 438
 FT DOMAIN 454 704
 FT ACT_SITE 495 495
 FT ACT_SITE 557 557
 FT ACT_SITE 651 651
 FT MOD_RES 164 164
 FT DISULFID 78 96
 FT DISULFID 148 162
 FT DISULFID 158 171
 FT DISULFID 173 186
 FT DISULFID 190 217
 FT DISULFID 247 265
 FT DISULFID 306 354
 FT DISULFID 334 367
 FT DISULFID 372 419
 FT DISULFID 402 437
 FT DISULFID 441 577
 FT DISULFID 619 636
 FT DISULFID 647 677
 FT CARBOHYD 54 54
 FT CARBOHYD 183 183
 FT CARBOHYD 390 390
 FT CARBOHYD 412 412
 SO SEQUENCE 704 AA; 79895 MW; 71F44F3012D2C67F CRC64;
 Query Match 12.5%; Score 586.5; DB 1; Length 704;
 Best Local Similarity 25.3%; Pred. No. 2.7e-32;
 Matches 191; Conservative 107; Mismatches 261; Indels 197; Gaps 31;
 QY 220 ARGELM-----RTTPGPPDSPPAHARQWALRGDADSVSLTFRSFDLASCDCERGSDL 275
 DB 24 AHYELNEMHQIQISPGPDS-YPSDSVETWNITVPEGFRIKLTFMHRNLSVLCEDY 82
 QY 276 VTIVNLTSPMPHALVOLCGTTPPSYNLTFFHSSQNVLL-----ITLIT--INERRAPG 326
 DB 83 VKV-----ETEDOVLATFCGRRTTDEOT--PCGOEVLVSPCTFMSVFRSDFSNERRFTG 135


```

FT DISULFID 143 157 POTENTIAL
FT DISULFID 153 166 POTENTIAL
FT DISULFID 168 181 POTENTIAL
FT DISULFID 185 212 POTENTIAL
FT DISULFID 242 260 POTENTIAL
FT DISULFID 301 349 POTENTIAL
FT DISULFID 329 362 POTENTIAL
FT DISULFID 367 414 POTENTIAL
FT DISULFID 397 432 POTENTIAL
FT DISULFID 436 572 INTERCHAIN (POTENTIAL)
FT DISULFID 614 631 POTENTIAL
FT DISULFID 642 672 POTENTIAL
FT CARBOHYD 49 49 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 178 178 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 385 385 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CARBOHYD 407 407 N-LINKED (GLCNAC. .) (POTENTIAL)
FT CONFLICT 235 235 E -> Q (IN REF. 2)
FT CONFLICT 285 285 G -> A (IN REF. 2)
FT CONFLICT 499 499 K -> Q (IN REF. 2)
FT CONFLICT 543 543 K -> Q (IN REF. 2)
FT CONFLICT 643 643 S -> A (IN REF. 2)
SEQUENCE 699 AA: 69261 MW: 45DEC2EA6EB40151 CRC64;

```

Query Match 12.38; Score 577; DB 1; Length 699;
 Best Local Similarity 25.48; Pred. NO. 1.2e-31;
 Matches 194; Conservative 110; Mismatches 268; Indels 192; Gaps 32;

```

QY 203 SKTQRTQDNDSCSGFGLHARGVELRFTTGPDPSPYPARHARCOMALRGDADSVLTFRS 262
DB 15 SKASAHTELVNMMG-----QIOSPGTPDS-YPSDEVNNTVYVPGCFRIKLFYMH 64
QY 263 FDLASCDERGSGLTYVNTLSPMEPHALVOLCG-----TYPPSY-NLTFH 306
DB 65 FNLSSYLCEVDYKV-----ETEDQYLATFCGRETDTDEGTGQGEVLVSGSPMSITFR 119
QY 307 SSONVLLTLTTERHHPGFEATFOLP----- 335
DB 120 SD-----FSNEERFTGDAAHYMAVDECKEREDELSCDHCHNYIGGYCCSRE 170
QY 336 -----RMSSCGRLKAKAGTFNSPYPGHYPPNIDCTWNIENYNNQHKVSKFE 384
DB 171 GYLHNTNRICRVCSNLFQRTGVITSPFPPRYKSSCLTYLEEEGFMYNLOPED 230
QY 385 FYLLE-PGVAGTCKPKDYVEIN-GEK-----YCGERSQFVNTSNKITYRHSDOYTD 437
DB 231 IFDIEDHPEV--CPDYIKIKVGPVILGPFCEKAPRPISTQSHSLVLFHSDNGEN 287
QY 438 TGFIAEYLSDDSPCP-----GQTCRTGRCIRKELRCDGMADCTHSDDELNCSCDA 490
DB 288 RGMRLSYRA--AGNECEPELQRPVHGKIETPSQAKYFFK-----DOYLVSCDT 331
QY 491 GHOF-----TCNKKFCKPLFW-----VDSVNDGCGNDSDEGCSCPAQ-----TFR 531
DB 332 GYKVLKDNVEMDTQIECLDGTWSNKIPTCKIYD-----CARPGLHGLITFS 381
QY 532 CSNKGCLSKSQCGKDGDCGDSDEASCPRVYVTCRKHYRC-----INGLCLSKGNP 585
DB 382 TRNMLTTYKSEI---KYSCE-----PYKMLNNTGITYTCSAOGVMWVKV-LGRSLP 430
QY 586 ECDDEKESDSDKDDCCGARTRO--ARVAGCTDADEEMWOWSLHAGGHICGA 643
DB 431 TGLPV-----CGLPKFSRLMARIFNGRPAOKTTPWIMLSLHNLQAPFCG 477
QY 644 SLISPMNLVSAHCV---ID--DRGFRSD--PTQMTAFGLDQSORSAVGOERRLK 695
DB 478 SLGSSMIVTAHRLHSLQPKDPTLSDLSLSPDKIILGKH-WRLRSSENGHGLGV 536
QY 696 RIISHPEFNDFTEDYDIALLELEKPAEYSSMWRPCLPDASHVPAKAIWVTGMGHTOY 755
DB 537 HTLHPKPYDPTFENDVALVELLESPLNAFVMPICLPEGPQ--QEGAMVIVSGWG-KOF 593
QY 756 GGTGALLLQGEIYVNOTJCENL---LPOQITRPMACVGLSGVDSGCGDGLPLSV 812

```

```

DB 594 LORPELTMELEIPIVHSTOCKAYAPLKKRVTRDMICAGEKEGKDACSDSGCPWVL 653
QY 813 EAD-GRIFQAGVSMGDCAOBNKPGVYTRLPFRDMIKENTGV 855
DB 654 NRENGOMLVGTVSMGDDCGKKDKRGVYSTIHNKMDIQRVTGV 697

```

```

RESULT 12
TMS2 HUMAN
ID TMS2 HUMAN STANDARD; PRT; 492 AA.
AC 015353;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRANSMEMBRANE PROTEASE, SERINE 2 (EC 3.4.21.-).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
ON NCBI-Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97468144; PubMed-9325052;
RA Peoloni-Giacobino A., Chen H., Pelisch M.C., Rossier C.,
RA Antonarakis S.E.;
RT "Cloning of the TMPRSS2 gene, which encodes a novel serine protease
RT with transmembrane, LDLRA, and SRCR domains and maps to 21q22.3.";
RL Genomics 44:309-320(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX PubMed-11414763;
RA Tang D.H., Chen Y., Lian L., Ha P.C., Tavtigian S.V., Wong A.K.;
RT "Mutation analyses of 268 candidate genes in human tumor cell lines.";
RL Genomics 74:352-364(2001).
RN [3]
RP TISSUE SPECIFICITY.
RA MEDLINE-21104370; PubMed-11169526;
RA Veerala M.H., Porvari K.S., Kellokumpu S., Kyllonen A.P., Vihko P.T.;
RT "Expression of transmembrane serine protease TMPRSS2 in mouse-and
RT human tissues.";
RL J. Pathol. 193:134-140(2001)
CC -|- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -|- TISSUE SPECIFICITY: EXPRESSED STRONGLY IN SMALL INTESTINE. ALSO
CC EXPRESSED IN PROSTATE, COLON, STOMACH, AND SALIVARY GLAND.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -|- SIMILARITY: CONTAINS 1 SRCR DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
DB EMBL: U75329; AAC51784.1;
DB EMBL: AF123453; AAD37117.1;
DB HSSP: P00763; IDPO.
DB MIM: 602060;
DB InterPro: IPR001314; Chymotrypsin.
DB InterPro: IPR002172; LDL_receptL_A.
DB InterPro: IPR001190; SRCR.
DB InterPro: IPR001254; Trypsin.
DB Pfam: PF00057; IdL_receptL_A; 1.
DB Pfam: PF00089; trypsin; 1.
DB PRINTS: PR00722; CHYMOTRYPSIN.
DB SMART: SM00192; IDLA; 1.
DB SMART: SM00202; SR; 1.
DB SMART: SM00202; TRYP; 1.
DB PROSITE: PS01209; LDLRA_1; 1.

```

DR PROSITE: PS00068: LDLRA_2: 1;
 DR PROSITE: PS00420: SRCR_1: FALSE_NEG;
 DR PROSITE: PS50287: SRCR_2: 1;
 DR PROSITE: PS50240: TRYPSIN_DOM: 1;
 DR PROSITE: PS00134: TRYPSIN_HIS: 1;
 DR PROSITE: PS00135: TRYPSIN_SER: 1;
 KW Hydrolase; Serine protease; Transmembrane; Signal-anchor.
 FT DOMAIN 1 84
 FT TRANSMEM 85 105
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT CYTOPLASMIC (POTENTIAL)
 FT EXTRACELLULAR (POTENTIAL)
 FT LDL-RECEPTOR CLASS A.
 FT SRCR.
 FT SERINE PROTEASE.
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CLEAVAGE (POTENTIAL).
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT BY SIMILARITY.
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT L-> I (IN REF. 2).
 FT Q-> E (IN REF. 2).
 FT N-> K (IN REF. 2).
 FT KAN -> RAD (IN REF. 2).
 SQ SEQUENCE 492 AA: 53847 MW: 78AAFFDAJ18609DDA CRC64:
 Query Match 12.0%; Score 560.5; DB 1; Length 492;
 Best Local Similarity 33.9%; Pred. No. 1e-30;
 Matches 131; Conservative 58; Mismatches 128; Indels 69; Gaps 14:
 QY 522 GCSPPAOTFRG-SNGKQLSKSQCCNGKDCGDSDEASCPKV---NVV----- 565
 DB 110 GSKCSNGTECDSSGTCINPSNWCDSVSHCPGDEENRCVRLYGFNFILQWYSSORKSWH 169
 QY 566 -----TCT---KHTRCLNG-----LCLSKNPECDCGKEDCDSD 595
 DB 170 PYCDDDMNENYGRAACRDGMGYNKFYSQGIYVDDSGSTSFMKLNTSAGNVDIYKLYHSD 229
 QY 596 GSEK-----DCDGLRSTFROAVVGGTDADEGWPMQVSLHALGQGHICCASLIS 647
 DB 730 ACSSKAVSLRCLAGVNLNS-SROSRIVGESALPGAWPMQVSLH-VONVHVCGSIIIT 287
 QY 648 PNWLSAAHCHYIDDRGFRYSPTQWTAFLGLHDQS-QRSAPGVQERRLRRIISHPFNFND 766
 DB 288 PEMIVTAHCHVEKP---LNNPWHHTATAGILRQSFMEYAGCY---VQKVISHPNYDSK 340
 QY 707 TFDYDIALLEKPAEYSNWRPCLPDASHVFPAGKAIWVTGWGHTQYGGTGAILQKQ 766
 DB 341 TKNNIDIALMKLOKPLTFNDLKVPCVLPNPGMMLQPEQLCWISGWCATEEKGKTSVLNAA 400
 QY 767 ELRVNQTTCEN-LLPQOITPRMCMVGLSGVSDSCGDSGLSSVEADGRI-FQAGV 823
 DB 401 KYLLIETQRCNRYRYNDLITPAMICAGFLQGNVDSCGDSGGPL-VTSNNNNWNLIGD 458
 QY 824 VSWGDCGCAQRNPGVYVTLPLFRDWI 849
 DB 459 TSMGSCCAKAYRPGVYGNVMTDNI 484
 RESULT 13
 KAL_MOUSE
 ID KAL_MOUSE
 AC P26262; STANDARD: PRT: 638 AA.

DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PLASMA KALLIKREIN PRECURSOR (EC 3.4.21.34) (PLASMA PREKALLIKREIN)
 DE (KININOGENIN) (FLETCHER FACTOR).
 GN KLK3 OR PK.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID:10090;
 RN [1]
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN-BALB/C; TISSUE-Liver;
 RX MEDLINE:91090844; PubMed:2264928;
 RA Seidah N.G., Sawyer N., Hamelin J., Mion P., Beaulieu G.,
 RA Brachpala L., Rochemont J., Melikay M., Chretien M.,
 RT "Mouse plasma kallikrein: cDNA structure, enzyme characterization,
 RT and comparison of protein and mRNA levels among species";
 RL DNA Cell Biol. 9:737-748(1990).
 CC -!- FUNCTION: THE ENZYME CLEAVES LYS-ARG AND ARG-SER BONDS. IT
 CC ACTIVATES, IN A RECIPROCAL REACTION, FACTOR XII AFTER ITS BINDING
 CC TO A NEGATIVELY CHARGED SURFACE. IT ALSO RELEASES BRADYKININ FROM
 CC HMW KININOGEN AND MAY ALSO PLAY A ROLE IN THE RENIN-ANGIOTENSIN
 CC SYSTEM BY CONVERTING PRORENIN INTO RENIN.
 CC -!- SUBUNIT: THE ZMOGEN IS ACTIVATED BY FACTOR XIIA, WHICH CLEAVES
 CC THE MOLECULE INTO A LIGHT CHAIN, WHICH CONTAINS THE ACTIVE SITE,
 CC AND A HEAVY CHAIN, WHICH ASSOCIATES WITH HMW KININOGEN. THESE
 CC CHAINS ARE LINKED BY ONE OR MORE DISULFIDE BONDS.
 CC -!- SIMILARITY: CONTAINS 4 APPLE DOMAINS.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. PLASMA KALLIKREIN SUBFAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL: M58588; AAA63393.1;
 DR PIR: A36557; KOWSPL.
 DR HSSP: P00750; IRTF.
 DR MEROPS: S01.212;
 DR MGD: MGI:102849; KLK3.
 DR InterPro: IPR000177; Apple.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR003014; PAN.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00024; PAN; 4.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00005; APPLEDOMAIN.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00223; APPLE; 4.
 DR SMART: SM00020; Tryp_SPC; 1.
 DR PROSITE: PS00495; APPLE; 4.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Plasma; Zymogen; Signal;
 KW Fibrinolysis; Blood coagulation; Inflammatory response; Liver;
 KW Repeat.
 FT SIGNAL 1 19
 FT CHAIN 20 390
 FT CHAIN 391 638
 FT DOMAIN 20 105
 FT DOMAIN 110 195
 FT DOMAIN 200 285
 FT DOMAIN 291 376
 FT DOMAIN 389 621
 FT CARBOHYD 127 127
 FT CARBOHYD 308 308
 FT CARBOHYD 396 396
 PLASMA KALLIKREIN HEAVY CHAIN.
 PLASMA KALLIKREIN LIGHT CHAIN.
 APPLE 1.
 APPLE 2.
 APPLE 3.
 APPLE 4.
 SERINE PROTEASE.
 N-LINKED (GLCNAC. . .) (PROBABLE).
 N-LINKED (GLCNAC. . .) (PROBABLE).
 N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 453 453 O-LINKED (PROBABLE)
 FT CARBOHYD 494 494 N-LINKED (GLCNAC...) (PROBABLE)
 FT ACT_SITE 434 434 CHARGE RELAY SYSTEM.
 FT ACT_SITE 483 483 CHARGE RELAY SYSTEM.
 FT ACT_SITE 578 578 CHARGE RELAY SYSTEM.
 FT DISULFID 21 104 BY SIMILARITY.
 FT DISULFID 47 77 BY SIMILARITY.
 FT DISULFID 51 57 BY SIMILARITY.
 FT DISULFID 111 194 BY SIMILARITY.
 FT DISULFID 137 166 BY SIMILARITY.
 FT DISULFID 141 147 BY SIMILARITY.
 FT DISULFID 201 284 BY SIMILARITY.
 FT DISULFID 227 256 BY SIMILARITY.
 FT DISULFID 231 237 BY SIMILARITY.
 FT DISULFID 292 375 BY SIMILARITY.
 FT DISULFID 318 347 BY SIMILARITY.
 FT DISULFID 322 328 BY SIMILARITY.
 FT DISULFID 340 345 BY SIMILARITY.
 FT DISULFID 383 503 BY SIMILARITY.
 FT DISULFID 419 455 BY SIMILARITY.
 FT DISULFID 517 584 BY SIMILARITY.
 FT DISULFID 548 563 BY SIMILARITY.
 FT DISULFID 574 602 BY SIMILARITY.
 SQ SEQUENCE 638 AA: 71368 MW: CC27C93AB1086599 CRC64:

Query Match 11.48; Score 533.5; DB 1; Length 638;
 Best Local Similarity 27.18; Pred. No. 9.2e-29;
 Matches 183; Conservative 86; Mismatches 223; Indels 173; Gaps 31;

QY 271 RGSLLVYVNTLSMEHALVOLCGTTPSYNLTFHSSONVLLTLTTERHPGF-E 328
 DB 32 RGGGLARIYTP---DAOYCKMCTFHPRLCLFSF-----LAVTPKRETKRKRCCEMKE 81
 QY 329 ATEPOLRMSSCG---GLRKAOGTFNSPYRPHYPP-----NDICTWNIE-----V 372
 DB 82 SINTGLRIHRTGAISSHLSKCOCHOISACHRDLYKGLDMGSMFNISKDNIECKKLC 141
 QY 373 PNNQHV---VSEKFEYLL-----PGVPAGT-----C 397
 DB 142 TNNHCOFFYATSAFYRPEYRKCKLKHSAAGTPTSIKSADNLVSGFSLSKALSEIGC 201
 QY 398 PKD-----VEINGEKNGGERS---QFVYTSNNKITVRHSSOSTDTC---FLAEY 444
 DB 202 PMDFHSAFADLNVSOVITPDPAVCRITCFHPNCLFETTYTNEWETESORNVCFKTS 261
 QY 445 LSYDSDPCFQO-----FTCRTR---CIRKELRCDGADCTDHSDELNCSGAGH- 492
 DB 262 KSGRPSPRIQENALISGTSLLTCKRTRECHSKI-----YSGVDFEGELNVTFFVGAD 316
 QY 493 --OPTCKNKKCKPLFWQSY-NDGGNSDEOGSCPAOTFRCSNGKCLSKSQCCNGKDD 549
 DB 317 VCOETC--TKTRICQFFIYSLPDC-----KEEGCKC---SLRLST----- 353
 QY 550 CGDSDSASCPKVVVCTKHTYRCLNG-----LCLSKGPECDGKEDCSGSDSDKDD 603
 DB 354 --DQSP-----TRTIGMOSSGYSRLCKLVDSPTTKYN----- 388
 QY 604 CGLSFTROARVVGTDADDEGPMQVSLNA--LGQGHICASLSPNWLVSAAHCYIDD 661
 DB 389 -----ARIVGTNAGISEMPQVSLQKLVSOYTHLGGSLIGQWVLTAAHCF--- 436
 QY 662 RGFYSDPTOWTALGLHDSO--RSAPGOERLKRIRISFPFNDTFYDIALLELEK 719
 DB 437 DGIRY--PDWIRIYGLISLSEIKETP--SSRIKEILIHQEKVSEGNDAIALIKLOT 491
 QY 720 PARYSSMWPRICLPDASHVFPAGKAIWVGTHOYGTGALLIOKGEIRVINTGTENL 779
 DB 492 PLNTFFOKPCLPSKADNTIYTNCHVGTGKYEKGEGEONITLOKATIPLVNPEECCOK 551
 QY 780 LPQO--ITPRMVCVGLSGVDSGCGSGSLSYEADGRIFQAGVNSKGDCAQRNPKGV 838
 DB 552 YRDVYINKMVICAGYKKGKDGKSGGRL--VCKHSGRMQLVGLTMSGEGCGKRDQGV 610

QY 839 YTRLPLFRDWIKENT 853
 DB 611 YTKSEYMDILEKT 625
 RESULT 14
 TMS3_HUMAN
 ID TMS3_HUMAN STANDARD; PRT; 454 AA.
 AC P57727;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE TRANSMEMBRANE PROTEASE, SERINE 3 (EC 3.4.21.-) (SERINE PROTEASE
 TADG-12) (TUMOR ASSOCIATED DIFFERENTIALLY EXPRESSED GENE-12 PROTEIN).
 GN TMRPS3 OR TADG12 OR ECHOS1.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eulheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_Taxid:9606;
 RP [1]
 RC TISSUE-Ovarian carcinoma;
 RX MEDLINE-20521358; PubMed-11068177;
 RA Underwood L.J., Shigenaga K., Tanimoto H., Beard J.B., Schneider E.N.,
 RA Wang Y., Parmley T.H., O'Brien T.J.;
 RT "Ovarian tumor cells express a novel, multi-domain cell surface serine
 RT protease.";
 RL Biochim. Biophys. Acta 1502:337-350(2000).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A, B/C AND D), AND VARIANT ILE-53.
 RX MEDLINE-20578749; PubMed-11137999;
 RA Scott H.S., Kudoh J., Wattenhofer M., Shibuya K., Berry A., Chirasil R.,
 RA Giupponi M., Wang J., Kawasaki K., Asakawa S., Minoshima S.,
 RA Younus F., Mehdi S.O., Radhakrishna U., Papasavas M.P., Gehrig C.,
 RA Rossier C., Korostishevsky M., Gal A., Shimizu N., Bonne-Tamit B.,
 RA Antonarakis S.E.;
 RT "Insertion of beta-satellite repeats identifies a transmembrane
 RT protease causing both congenital and childhood onset autosomal
 RT recessive deafness.";
 RL Nat. Genet. 27:59-63(2001).
 CC - FUNCTION: PROBABLE PROTEASE.
 CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
 CC - ALTERNATIVE PRODUCTS: 4 ISOFORMS; A (SHOWN HERE), B/C, D AND
 CC TRUNCATED/TADG-12V; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC - TISSUE SPECIFICITY: EXPRESSED IN MANY TISSUES. THE TRUNCATED
 CC ISOFORM IS FOUND AT INCREASED LEVELS IN SOME CARCINOMAS.
 CC - DISEASE: DEFECTS IN TMRPS3 ARE A CAUSE OF TWO FORMS OF AUTOSOMAL,
 CC NEUROSENSORY CHILDHOOD-ONSET FORMS OF DEAFNESS, DFN8 AND DFN10.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC - SIMILARITY: CONTAINS 1 SRCR DOMAIN.
 CC - SIMILARITY: CONTAINS 1 LDL-RECEPTOR CLASS A DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

DR EMBL: AF201380; AAC37012.1;
 DR EMBL: AB038157; BAB20077.1;
 DR EMBL: AB038158; BAB20078.1;
 DR EMBL: AB038159; BAB20079.1;
 DR EMBL: AB038160; BAB20080.1;
 DR MIM: 605511;
 DR MIM: 601072;
 DR MIM: 605316;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR002172; LDL_recept_A.
 DR InterPro: IPR001190; SRCR.

